Payment Systems and Network Effects

ISBN 90-9018112-1

Cover design: Fridy Visser

Payment Systems and Network Effects

Adoption, Harmonization and Succession of Network Technologies Across Countries

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen op donderdag 3 juni 2004 om 16.00 uur

door

Johan Gottfried Leibbrandt

Promotor: Prof. dr. R. Cowan

Beoordelingscommissie: Prof. dr. P.A. Mohnen (voorzitter)
Prof. dr. P.A. David (Stanford University)
Prof. dr. S. Kleimeier

"Alas! How deeply painful is all payment"
Lord Byron

Contents

Preface **xi**

1 Facts and Questions **1**
1.1 Main Payment Instruments 1
1.2 Usage patterns across countries 4
1.2.1 Usage of cash 4
1.2.2 Usage of non-cash instruments 5
1.3 Evolution and succession of non-cash instruments 9
1.3.1 US: Darwin without the extinction of the dinosaurs . . 9
1.3.2 Netherlands: four weddings and a funeral 10
1.3.3 Comparison of payment instrument succession in US and Netherlands 12
1.4 Economic Relevance 13
1.5 Questions for this thesis 15

2 Theory of payment instruments and networks **17**
2.1 Payment instruments 17
2.1.1 Cash versus alternatives 18
2.1.2 Will cash disappear? 20
2.1.3 Choice between non-cash instruments 21
2.1.4 Explaining country differences 25
2.2 Network externalities 26
2.2.1 Unsponsored standards 27
2.2.2 Sponsored standards 30
2.2.3 Role of autarky and spatially separated users 33
2.3 Payment Instruments as Networks 34
2.3.1 Empirical evidence 34
2.3.2 Theoretical models of payment instruments as networks 36
2.3.3 Regulatory implications 38
2.4 Conclusions from payment and network literature 39

3 Adoption and harmonization of unsponsored standards **41**
3.1 Basic model 42
3.2 The adoption decision 44
3.2.1 Concept of critical share and role of industry structure . 44
3.2.2 Effect of upgrading an old technology 48
3.3 Autarky and the adoption decision 49

3.3.1 Autarkic banks 50
3.3.2 Multiple players in semi-autarkic countries 52
3.4 The compatibility decision 54
3.5 Summary and conclusions 58

4 Adoption and harmonization of sponsored standards 61
4.1 The basic duopoly model 62
4.1.1 Base case ($\varepsilon = 0, \delta = 1, b < t$) 64
4.1.2 Semi-autarkic transaction patterns ($\delta < 1$) 69
4.1.3 Strong network externalities: $b \geq t$ 72
4.1.4 Variable transaction demand ($\varepsilon > 0$) 73
4.2 Adoption of sponsored standards by an oligopoly 77
4.2.1 A model of competition by an asymmetric oligopoly . . 77
4.2.2 Possible equilibria 79
4.2.3 Analysis of four polar market structures 81
4.3 Discussion of results 86
4.3.1 Comparison of results for sponsored standards with unsponsored standards 86
4.3.2 Welfare effects 87
4.4 Discussion of model and main assumptions 89
4.4.1 Does Nash-equilibrium apply? 89
4.4.2 Comparison to the model of Shy (2001) 91
4.5 Conclusions 93
4.6 Chapter appendix: symbols used in chapters 3 and 4 96

5 Case 1: adoption of giro-systems 97
5.1 Case background and methodology 97
5.2 How giro was introduced 98
5.3 The economics of giro-systems 102
5.4 Applying the model 105
5.4.1 Estimating parameters for the unsponsored model . . . 105
5.4.2 Estimating parameters for the sponsored model 107
5.4.3 Applying the model for sponsored standards 109
5.5 Discussion of results 110

6 Case 2: European harmonization of ACH systems 113
6.1 Case background and method 113
6.2 Creating a Single Euro Payments Area (SEPA): description of events 114
6.2.1 1990s: "Europe" grows increasingly frustrated with cross-border transfers 114

6.2.2 2001: legislation on pricing of Euro payments 116
6.2.3 2002: the SEPA workshop 117
6.2.4 Other initiatives 120
6.3 European ACH/giro landscape 120
6.3.1 Incompatible systems 120
6.3.2 Autarkic countries 122
6.3.3 Industry structure: locally concentrated, fragmented at European level 123
6.4 Applying the model 126
6.4.1 Model for unsponsored standards 126
6.4.2 Model for sponsored standards 127
6.5 Discussion of results 128

7 Technical change and technology succession: introduction and theory 129
7.1 Innovation and technical change 130
7.2 Technology and country differences 133
7.3 Models of technology succession 135

8 A model for succession of payment networks 137
8.1 Description of the model 138
8.1.1 Basic elements 138
8.1.2 Incremental profit and the role of the installed base . . 140
8.1.3 Definition of equilibrium and regret 142
8.2 Analysis of number of equilibrium points and size of regret . . 145
8.2.1 Analysis of 2 by 2 case 145
8.2.2 Extension to larger m and n 147
8.2.3 Effect of cost and benefit structure on occurrence of multiple equilibria 148
8.3 Impact of initial differences in installed base 150
8.3.1 Mechanics of the simulation model and sample run . . . 152
8.3.2 Convergence and divergence across countries 156
8.3.3 Sensitivity to main parameters 159
8.3.4 Impact of economic factors on country convergence . . . 161
8.3.5 Effect of late starts 163
8.4 Incremental change versus paradigm shifts 164
8.5 Discussion of results 167
8.5.1 Is the model realistic? 167
8.5.2 Do the outcomes have meaningful implications? 168
8.6 Chapter appendix: list of symbols used 170

9 Case 3: Electronic money, Internet and Mobile payments **173**
9.1 The facts: what happened 173
9.1.1 Description of main contenders 173
9.1.2 Taking stock anno 2003 177
9.2 Existing literature and theory on electronic money and m-/e-payments 180
9.3 Applying the model 182
9.4 Conclusions of Internet payment case 184

10 Conclusions **185**

11 Epilogue: "To a man with a hammer .. **189**

References **191**

A Proof of propositions in chapter 3 **203**

B Proof of propositions in chapter 4 **207**

C Proof of propositions in chapter 8 **225**

D Distribution of technical change sizes **229**

Summary in Dutch 231

Curriculum Vitae 235

List of Tables

1.1 Transaction volume and value of main payment instruments . . 2
1.2 Ben-David convergence test for BIS-11 countries 8
1.3 Cost per non-cash transaction 14

3.1 Summary of equilibria with unsponsored standards 59

4.1 Equilibrium profits for stage 2 of base case duopoly game . . . 67
4.2 Payoff matrix for stage 1 of base case duopoly game 68
4.3 Payoff matrix with compatibility ex-post 69
4.4 Payoff matrix with 3 equal sized firms and $b = 0.8$ and $c = 0.12$ 80
4.5 Comparison with results for unsponsored standards 87
4.6 Welfare effects of various equilibrium outcomes 88
4.7 Minimax outcomes for a symmetric duopoly 91
4.8 Summary of Shy's (2001) results 92
4.9 Results of Shy's approach using my parameters 92

5.1 Giro-systems in Europe 99
5.2 Key data on costs and output for the Dutch giro-system 104
5.3 Critical share for giro adoption: sensitivity to cost assumptions 106
5.4 Share of branches for Dutch banks, 1966 108

6.1 European domestic account formats 121
6.2 Share of foreign transactions by country 122
6.3 Bank concentration in major markets, 2000 124
6.4 Number of payment networks by country 126

8.1 Benefit and cost components of ATM and POS technologies . . 139
8.2 Average results for 2 countries across 10,000 runs of model . . . 156
8.3 Different endpoints for 10 countries with initial differences . . . 159

9.1 Electronic Internet payment methods in EU and US 178
9.2 Usage of existing infrastructure by new payment schemes . . . 183

D.1 Results of fitting distribution to increases in costs 229

List of Figures

1.1 Usage of non-cash instruments in 11 BIS countries 6
1.2 Evolution of payment instruments in the US 11
1.3 Evolution of payment instruments in the Netherlands 11

3.1 Lock-in as a function of cost/benefit ratio and industry structure 46
3.2 Occurrence of lock-in with upgrades 49
3.3 Critical share with semi-autarky: $\delta = 0.5$ 52

4.1 Equilibria for semi-autarkic transaction patterns ($\delta < 1$) 71
4.2 Equilibria for large network effects ($b \geq 1$) 73
4.3 Equilibria for price sensitive demand ($\varepsilon > 0$) 75
4.4 Equilibrium outcomes for Gorilla vs fringe 84
4.5 Equilibrium outcomes in a fragmented market 85
4.6 Relative loss of social welfare with unsponsored standards . . . 89

5.1 Cost of a giro transaction 103

8.1 Probability of multiple equilibria as a function of m and n . . . 149
8.2 Average potential regret as a function of m and n 149
8.3 Impact of minimal changes in cost matrix on equilibrium points 151
8.4 Stand alone profit of technologies in sample run of model . . . 155
8.5 Technology paths for 2 counties in sample run of model 157
8.6 Development of the installed base for 2 countries in sample run 157
8.7 Convergence as a function of m and n 160
8.8 Average regret as a function of m and n 160
8.9 Convergence as a function of parameter p 161
8.10 Average regret as a function of parameter p 162
8.11 Convergence as a function of global scale economies 163
8.12 Convergence as a function of the country 2 adoption delay . . . 164
8.13 Country 2 profit advantage as a function of adoption delay . . 165
8.14 Distribution of increase in benefit components after each adoption 166
8.15 Distribution of profit increase after each adoption 166

C.1 Fit of predicted probability of multiple equilibria 228

D.1 Incremental costs per adoption for $m = 50$. 230
D.2 Incremental profits per adoption for $m = 50$. 230

Acknowledgements

First and foremost I would like to thank my sponsor Robin Cowan. Throughout the process he has been both supportive and critical, providing input where needed and staying out of the way when not. I now know a little more about network theory and I think he has learned something about payment systems. I am indebted to the members of the reading committee, Pierre Mohnen, Paul David and Stefanie Kleimeier, for their willingness to read the draft, including the appendices.

I would like to thank Luc Soete and the others at Merit, and particularly Wilma Coenegrachts and Silvana deSanctis for providing support to someone who's main contribution was asking ignorant questions during the Tuesday seminars. Müge Özman and Elad Harison provided very pleasant and helpful company during my days at Merit, helped me with academic software, and provided shelter in Maastricht when hotels were full.

Arnout Boot, Jean Tirole and J.-J. Rochet helped me find my way in the wonderful world of academia, pointing out such elementary things that most literature can be obtained by walking into the university library. Martin Fase set me up to present at the SUERF conference in Brussels in 2001, providing a deadline and platform that greatly accelerated my work.

David Humphrey gave comments on early drafts, patiently pointed out his low tolerance for contrived theories and models (which, I think, included most network stuff) and provided me with several very relevant working papers on the costs of payments.

Simon Lelieveldt has been very helpful in educating me on the history of Dutch payments, in making available his substantial archive on the matter, and in providing detailed comments on the relevant chapters. Ruud den Hollander provided early data on the PCGD, while Gerard Hartsink and Robert Heisterborg provided valuable comments on the SEPA workshop and its follow-up.

My colleagues at McKinsey are of course to be thanked for their support throughout the journey. In particular Olivier Denecker for providing me with the data on European banking concentration and cross-border card usage, Robert Reibestein for his impatience when I failed to finish my PhD within the 2 year timeframe he thought reasonable, Pieter Winsemius for the idea of making a layman's version, and Alexine Mulock Houwer for proofing, printing and moral support.

Several people helped me getting the text ready for printing. Maarten Pieter Schinkel helped me with the black art of Latex and Fridy Visser designed the cover.

And finally of course there is my family. Maartje read and proofed part of the draft, Jan , Kick and Lotte listened to stories about credit cards and all of them motivated me by at least creating the impression that a dad/husband with a PhD is cool (as long as he does not wear white socks or reading glasses).

Preface

In a world obsessed with globalization (be it in favor or against), it is perhaps interesting to note that many things are still local. Consider the following list:[1]

110 Volt	220 Volt
English	French
Common law	Continental law
CDMA	GSM
NTSC	PAL
Folio	A4
3-ring binders	2- or 4-ring binders
QWERTY	AZERTY
British imperial units	Metric
80 proof	40%
Left hand drive	Right hand drive
Personal checks	Giro transfers
Signature	PIN

All the items are examples of standards, de facto or de jure. The last two items are taken from the payments industry, and appear to be the out of place on this list. The central theme of this thesis is that they *should* be on the list. Payment systems are subject to the same economic forces that lead different standards to persist across countries.

The past century has seen a rapid evolution in the use of payment instruments. While cash and the occasional check dominated the landscape in 1900 this picture had changed dramatically by 2000, with widespread use of transfers, direct debits and various cards. At first sight this represents a classic example of economical progress: innovation leads to new instruments that deliver more benefits at lower costs, and as they are adopted by economic agents these innovations displace older technologies. However, upon closer inspection things are not that simple. In particular three phenomena occur that run counter to this 'neoclassical' view:

1. Adoption differs across countries. Some countries fail to adopt an economically superior instrument.

[1] Loosely adapted from the January 2002 issue of Wired magazine, p89.

2. Countries adopt incompatible versions, and efforts to harmonize them frequently fail.

3. Initial differences in the technology base between countries appear to perpetuate themselves.

This thesis explores whether network externalities can help explain these phenomena. It is well-known that network externalities can lead to multiple equilibria; for example a particular standard is either adopted by everybody or by nobody. Network externalities could therefore explain the above three phenomena as follows. Because of the network externalities, a certain critical mass of banks is required to make the adoption of the newer payment technologies profitable. And because most transactions take place within rather than across countries, is the critical mass within a country that counts; this puts countries with an existing fragmented industry structure at a disadvantage in adopting such newer payment technologies. Due to that same autarky, the adopting countries often end up with incompatible versions. And finally, because of the network nature, the existing installed base influences which new technologies are adopted, perpetuating initial differences.

Precisely because this story sounds plausible, extreme care is required in assessing its veracity. Like the man in Maurits Escher's picture at an exhibition, network effects are subject to self-reference: they have been used to describe fads; but they themselves are also subject to the dynamics of fads. Network effects have been invoked to proclaim 'new rules for the new economy', a concept that has come crashing back to earth like a punctured hot air balloon. The purpose of this thesis is to provide the required rigor.

The overall proof rests on three pillars of evidence. The first is econometric analysis of network effects in specific payment instruments, most notably the extensive and recent work by Gautam Gowrisankaran and Joanna Stavins on US ACH credit transfers. The second is mathematical modelling of adoption, harmonization and succession of network technologies, building on the work of Oz Shy, Andre DePalma, Luc Leruth and many others. The modelling leads to some strong and robust results. This second pillar will occupy the bulk of this thesis. The third pillar is provided by the application of the models to three cases: ACH/giro adoption in the Netherlands, harmonization of European transfer payments, and the success of new payment technologies for mobile and Internet. As Tim Breshanan (2001) convincingly demonstrates in his paper "Network effects and Microsoft", documentary methods can provide a useful tool for testing theory, where econometric testing has proved difficult.

The results obtained in this thesis have implications far beyond payment systems. Whenever autarkic networks are involved, expect to see divergence

between the autarkic regions/countries; furthermore be prepared to recognize that such differences are very hard to harmonize; and that they can persist even if better technologies arrive on the (global) scene. In short, the world may be much more local than both neoclassical theory and many globalization gurus predict.

1
Facts and Questions

This first chapter describes the payment instrument landscape: what are the main instruments, how are they used and how did they evolve? I lay out the differences between countries: what instruments are used today, and how did the payments landscape evolve? This is followed by an analysis of the economic importance of differences in the usage of payment instruments; I find these differences to be economically significant. The last section of this introductory chapter formulates the questions to be addressed in this dissertation.

Throughout this chapter I rely on data from the Bank for International Settlements (BIS) in Basel. Their most recent 'red book' (BIS, 2003) contains figures on all major payment instruments for the years 1998-2002 for 11 countries: Belgium, Canada, France, Germany, Italy, Japan, Netherlands, Sweden, Switzerland, UK and US. I will refer to these as the BIS-11 countries.[1]

1.1 Main Payment Instruments

Following the classification of BIS, I distinguish seven main instruments which are described below.[2] Table 1.1 gives an overview of their usage in the BIS-11 countries.

1. *Cash.* The oldest and most widely used instrument. In terms of number of transactions, "cash is (still) king": the average person makes some 500 payments a year, and cash is used for the majority of these. There are relatively few sources of hard data on cash transactions; neither the number of transactions, nor their value is known with any certainty. Most data rely on samples and surveys.[3]

---

[1] Eight of these countries are European, and the data in chapter 6 (harmonization of EU transfer systems) will focus on these eight countries.

[2] See e.g. table 9 of the 2003 BIS "Red Book" (BIS 2003). While BIS data on payments are used throughout the literature, they are not without problems. The recent revision of US check use (it was revised downward by almost a third, see below) was spectacular, but there are other problems with the data; for an inventory see Norges Bank (2001).

[3] Surveys of cash usage are problematic, because people tend to underreport small purchases, see Alessie, Gradus, et al. (1990).

TABLE 1.1 Transaction volume and value of main instruments for BIS-11 countries, 2002

	Volume (blns)	Share of volume (%)	Value (EUR trlns)	Share of value (%)	Tx per person	Average tx value (EUR)
Cash	217.3	61%	1.3	1.0%	291	6
Checks	49.4	14	47.8	41.5	66	968
Credit trf.	21.5	6	58.3	42.4	29	2,710
Direct debit	14.5	4	15.3	12.5	19	1,055
Credit cards	22.0	6	1.8	1.7	29	86
Debit cards	24.6	8	1.2	1.0	39	39
E-purses	0.3	0	0.0	0.0	0	4
Total	354.4		125.7		474	355

Source: BIS, 2003 and De Grauwe, Buyst, et al., 2000.

2. *Checks.* This instrument is the most widely used non-cash instrument in the BIS-11 countries. This is largely driven by Americans who write 80% of all checks. Most of these are the familiar non-guaranteed checks written from a checkbook by consumers and businesses, although travellers checks, Eurocheques and bankers drafts are also included in this category. While easier to track than cash, data on checks are not as reliable as one would like. For example a recent study by the Federal Reserve System, (FRS, 2002) revised the number of checks written in the US in 2000 downward from 66 billion to 42.5 billion.[4]

3. *ACH credit transfers.* ACH (Automated Clearing House) credit transfers are used for many of the same transactions as checks, but unlike checks they are seldom used at the Point of Sale (POS). They differ from checks in that they are sent to the bank of the payor; this bank then executes the transaction either in-house (if the payee also has an account with the bank) or through a clearing house.[5]

[4] The uncertainty is due to the fact that most banks process 'on-us' checks in-house (with on-us checks the payee and the payor have an account with the same bank). These on-us checks include checks written by a customer on himself at the teller to obtain cash. Given the fragmentation of US banking, data on these on-us checks can only be captured through surveys.

[5] Technically this category also includes large value transfer systems that are mainly used for interbank payments, such as Fedwire and Target; these Real Time Gross Settlement (RTGS) systems directly post transactions to central bank accounts as they occur, providing immediate finality of payment.

4. *ACH direct debit.* In a direct debit the payor (debtor) has authorized the payee (creditor) to present the bank of the payor with an amount and account number to be credited. Direct debits are cleared through the same ACH's as the credit transfers, and in fact rely on much of the same technology.

5. *Credit Card payments.* These are payments using cards at the POS, where the money is debited to a 'card-account'. The resulting balance is presented monthly to the cardholder. If he has the option to pay only part of the balance, it is a true credit card, otherwise it is a charge card or a delayed debit card. This category includes the charge and credit cards of Visa, MasterCard, American Express, Diners etc., as well as many retailer cards (department stores, petrol chains).

6. *Debit Card payments.* These are card transactions that are directly debited to a demand deposit account. Two varieties of debit cards exist: (1) PIN-debit, where the customer uses the same PIN as for an ATM, and the transaction is routed through an ATM network (or something related), and (2) Signature debit; to a consumer and merchant these signature debit transactions are exactly like a Visa or MasterCard credit card transaction, except that they are directly debited to the consumer's checking account.[6]

7. *Card based e-purses and electronic cash.* Card based e-purses are loaded from the current account; this money can then be spent in stores, vending machines and over the phone or Internet. They have yet to gain wide usage.[7] A different type is e-cash, where the electronic 'bits and bytes' represent real money payable to bearer.[8] These bits can be stored on a chipcard or on a hard disk. Again, usage has been disappointing: the best-known examples, Mondex (cards) and eCash (hard disk), are both defunct.

The above list of instruments is neither complete nor unambiguous. It describes the main classes and species. It excludes for example the draft and bill

[6] For a detailed description see Caskey and Sellon (1994) and more recently Evans and Schmalensee (1999).

[7] To date the use has been limited. The only ones to be used by more than a handful of people are Proton in Belgium, and Chipknip in the Netherlands, which together recorded about 200 million transactions in 2002, or 8 transactions per capita per year (Source: websites of Interpay -interpay.nl- and Banksys -banksys.be). While low, the volume more than doubled compared to 2001; therefore, usage may still take off (see Lafferty Group, 2003).

[8] By contrast the e-purses of Proton and Chipknip do not contain real cash. Technically, a part of the current account is reserved for later usage.

of exchange, which are mainly of historical importance.[9] Also, new instruments keep appearing that are not easy to classify.[10]

1.2 Usage patterns across countries

It is difficult to obtain good data on cash, let alone data that can be compared across countries. Most of this thesis will therefore focus on usage of non-cash instruments. Nevertheless, I will first review the literature on the usage of cash, because cash is an important alternative for most non-cash payment instruments.

1.2.1 Usage of cash

Cash is the oldest and still most widely used instrument (in terms of volume of transactions). But surprising little is known about its use. Estimates of volume and average value of cash transactions are very rough at best. Since there is no systematic registration of cash transactions, we have to rely on surveys. These give widely varying results. A comprehensive overview of the available data and estimates is given by Boeschoten (1992).[11] He lists 20 estimates covering 12 countries for the use of cash; each estimate draws on a different survey and/or method. Estimates of the number of cash transactions per person per year vary from 86 (Japan, based on a 1990 study by the Bank of Japan) to 1120 (US, based on Humphrey and Berger, 1990). A recent estimate for Belgium is given in De Grauwe, Buyst, et al. (2000): 2.99 billion transactions (i.e. 291 per person), with an average value of EUR 6. I use this estimate in table 1.1, because it is both recent and in between the more extreme estimates mentioned earlier. All surveys find that cash is used for the majority of transactions, and that most transactions are extremely small. Using survey data from the Netherlands, Boeschoten and Fase (1989) provide evidence that the value of a cash transaction is lognormally distributed with an average value of $ 13 and

[9] A draft is essentially an IOU. A bill of exchange is a draft that has been endorsed by a bank. In combination with bilateral clearing between banks, these instruments were the main tools for trade payments (other than cash) from the renaissance to the early 20th century, when domestic trade switched to checks and giroclearing. Even today, most international payments are made through correspondent banking, which relies on banks maintaining mutual accounts and clearing any net balances through bills of exchange. For a description see Frankel and Marquardt (1987).

[10] For example ELV in Germany lets a consumer swipe his ATM (Eurocheque) card at the point of sale and sign a slip, after which the transaction is processed as a direct debit. It has taken BIS 5 years to figure out how to classify them. Initially registered as direct debits these payments are now (correctly) counted as POS debit card payments.

[11] Boeschoten (1992), table I-1, p. 200.

a modal value of $ 1.70.[12] Cash still makes up the majority of transactions in volume, but most of these are quite small. As a result cash represents only 1% in value terms.

It proves to be even more difficult to compare cash usage across countries. The most comprehensive cross-country comparisons have been made by Humphrey, Pulley, et al. (1996). They attempt to infer differences in cash usage from differences in currency stock (coins and notes) correcting for differences in GDP and population. As they themselves admit, this is problematic for two reasons. In the first place a large part of the supply of 'hard' currencies like the US dollar and the German mark (and presumably now the Euro) resides abroad, notably in South America and Eastern Europe/Russia.[13] Secondly, a disproportionate part of the currency stock is either 'hoarded' to avoid taxes or used in illegal activities.[14]

1.2.2 Usage of non-cash instruments

The growth in non-cash instruments has been spectacular: in the Netherlands the number of non-cash transactions went from about half a billion transactions in 1973 to 3 billion in 2001; this represents a growth of 6.5% per year over a 28 year period. Such growth of non-cash instruments occurs across all BIS-11 countries.[15] Using data over the years 1988-2001 (the only period for which consistent data across multiple countries are available), the average growth was 4.9% per year.[16] While this growth is fairly uniform, there are significant differences in the instrument mix across countries.

Figure 1.1 gives an overview of the mix of payment instruments for the BIS-11 countries. Three large differences stand out. First, the number of non-cash transactions per person per year varies a lot. It is very high in the US (270) and very low in Japan (28) and Italy (52); non-cash transactions per capita in the other eight countries are at an intermediate level of around 100 to 200 per

[12] Boeschoten and Hebbink (1996) state that similar patterns were found in surveys for Finland and France. The more recent estimate of the average transaction value by De Grauwe, Buyst, et al. (2000) is lower: EUR 6 vs. USD 13 found by Boescoten and Fase (1989). This could be explained by the fact that debit cards have taken over many of the larger cash transactions.

[13] Porter and Judson (1996) estimate that 60% of US currency is held outside the country. Seitz (1995) estimates that 30-40% of German currency is held abroad.

[14] Boeschoten (1992) estimates that across 14 OECD countries one-third of the value of currency in the hands of the public is used for such hoarding.

[15] This fact has been documented by, i.a. Hancock and Humphrey (1998).

[16] Surprisingly, most of the growth seems to have occurred in the late 1970s and early 1980s, before the advent of debit cards and other electronic instruments; non-cash transactions in the Netherlands grew by 5.3% per year from 1988-2001, compared to 7.4% in the years 1973-1987.

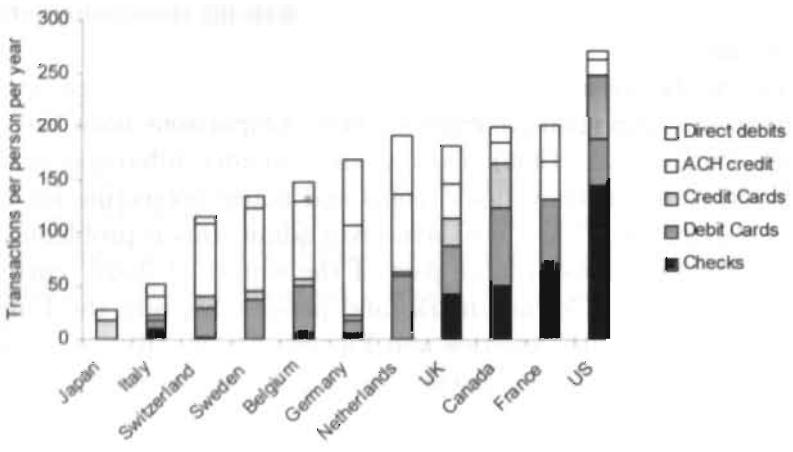

Figure 1.1 Usage of non-cash payment instruments across countries - 2002
Source: BIS, 2003.

year.[17] Second, the 'Anglo-Saxon countries' and France rely on checks, where the Benelux, Germany, Sweden and Switzerland use the ACH instruments of credit transfer and direct debit. Third, the Anglo-Saxon countries use credit cards extensively, while the others (except Italy) rely primarily on debit cards.

An important observation is that the use of payment instruments really follows national boundaries. A Frenchman just south of the Belgian border will write a check at the supermarket, while a Belgian just north of that border will use his debit card or e-purse. To be sure, there are variations in the use of instruments across the US: early debit card usage was concentrated in California and Pennsylvania.[18] Even today, 75% of the consumers in the West have a debit card compared to only 48% in the Midwest, while people in the Midwest write 27 percent more checks than those in the Northeast.[19] How-

[17] It seems strange that Japan has both a very low number of cash (86 per year) and non-cash transactions (28 per year) per capita. This could mean that the Japanese transact much less than others. A more plausible explanation is that the Bank of Japan study on cash transactions (on which the estimate of 86 cash transactions per capita per year is based) severely underestimates the real number of cash transactions; e.g. by not recording small cash transactions, which in other studies make up the bulk of cash transaction volume.

[18] Caskey and Sellon (1994).

[19] Debit figures from ABA (2001). Mantel and McHugh (2001) report a similar pattern. Check data from the 2002 check study (FRS, 2002) as summarized in Gerdes and Walton (2002).

ever, these are minor variations compared to the difference with Canada: the average American writes more than three times as many checks as the average Canadian.

Countries often use different, or at least incompatible versions of the same instrument. As an example, consider the giro-systems of the various Euro countries. It is easy to transfer money within each country, but much harder to do so across borders: the forms are different, the account numbers have different formats, and the rules (for resolving disputes, handling errors, etc.) vary by country.

These differences between countries cannot be fully explained by differences in maturity between markets. For example the US pioneered the credit card, but writes checks on a massive scale. In fact, rumors of the death of US checks are greatly exaggerated: the use of checks is at best stabilizing.[20] New electronic payment instruments may even have increased their use: "more checks are written to pay credit card bills, than are replaced at the point of sale" according to one estimate.[21]

Are these differences declining: is there convergence in terms of the use of payment instruments? To test for convergence, I apply a test proposed by Ben-David (1993) to measure convergence in per capita income across countries. It is applied here to non-cash transactions per capita. Ben-David models the following relation between per capita transactions relative to group average:

$$y_{i,t+1} - \overline{y}_{t+1} = \phi(y_{i,t} - \overline{y}_t)$$

where:

$$\begin{aligned} y_{i,t} &= \text{log per capita transactions of country } i \text{ in period } t \\ \overline{y}_t &= \text{unweighted average of } y_{i,t} \text{ across all countries in the group.} \end{aligned}$$

ϕ indicates declining ($\phi < 1$), increasing ($\phi > 1$) or stable ($\phi = 1$) differences in transactions per capita. I have estimated ϕ by regressing data on per capita

[20] There is some discussion as to whether their use is currently declining. Based upon recent studies by the Federal Reserve (FRS, 2002), Gerdes and Walton (2002) claim that their usage peaked in the early 1990s, and is currently about 5% below that peak level. However this claim appears dubious for two reasons: (1) the claim is based on comparing estimates of check use from three separate studies in 1979, 1995 and 2000, each of which had its own methodology; (2) the number of checks processed by the Federal Reserve (which is a hard figure) rose continuously from 16.2 billion in 1995 to 17.6 billion in 1999 and declined slightly to 17.5 in 2000. In fact Chakravorti and McHugh (2002) estimate that check usage grew by 22.7% from 1990 to 1999.

[21] See Murphy (1991).

TABLE 1.2 Results of Ben-David convergence test for BIS-11 countries, 1998-2002 ($\phi < 1$ indicates convergence, $\phi > 1$ means divergence)

Instrument	ϕ	Standard deviation	Half-life	Double-life
Checks	1.35	0.27		9
Credit transfers	0.94	0.04	46	
Direct debit	0.81	0.03	13	
Credit Cards	1.01	0.06	n/a	
Debit cards	0.74	0.02	9	
All non-cash instruments	1.03	0.04	n/a	

Source: data for 1998 and 2002 both from BIS (2003).

transactions for various instruments across the BIS-11 countries for $t+1 = 2002$ and $t = 1998$. The results are summarized in table 1.2.[22]

Overall, the picture is mixed: for three individual instrument there is convergence ($\phi < 1$), for checks there is divergence ($\phi = 1.35$), while for credit cards there is neither ($\phi \approx 1$). The overall number of non-cash transactions appears to be slightly diverging. With the exception of credit cards, all these values are significantly different from $\phi = 1$. To put the values of ϕ in perspective table 1.2 includes two columns 'half-life' and 'double-life', using the terms and definitions given by Ben-David. These represent the number of years needed to half the (log) difference with group average if $\phi < 1$ or double it if $\phi > 1$ (if $\phi \approx 1$ this is obviously not defined). Given the size of the differences in figure 1.1 it could take a while before these differences will disappear.[23]

It is interesting to this (lack of) convergence in payment instruments in the wider context of the convergence of banking markets. Convergence of banking markets has been high in the EU agenda, especially with the introduction of a single currency. Kleimeier and Sander have tested for convergence of European credit rates (K&S, 2002) and borrowing and lending spreads (K&S, 2000). They find only very limited (if any) evidence of such convergence. While rates have been converging, this has more to do with the single monetary policy for the Eurozone than with cross-border arbitrage of banks. In such cross-banking arbitrage the (threat of) entry by foreign players would force rates and spreads in various countries to move in line with each other. While they

[22] Japan was excluded from estimates on direct debits because the instrument is not used there. France was excluded from estimates on credit cards since the BIS does not distinguish credit and debit card transactions for that country.

[23] The results are the same if the sample is restricted to just the European countries, where one may expect more convergence due to the EURO, EU, 2nd banking directive etc. (see discussion further down in the main text). The value of ϕ for all non-cash payments goes from 0.93 for all countries to 0.92 for just the European countries.

find some evidence for cross-border arbitrage in corporate lending they find no evidence for this in consumer credit and mortgages. They conclude that (retail) banking markets remain national affairs. Different payment mechanisms may form a barrier for such cross-border arbitrage and therefor contribute to the lack of convergence in banking markets (and vice versa).

1.3 Evolution and succession of non-cash instruments

Not only the mix of instruments is different, but also the timing of their introduction and disappearance. As an example I will compare the recent history of two countries: the US and the Netherlands.[24]

1.3.1 US: Darwin without the extinction of the dinosaurs

Figure 1.2 gives a stylized overview of the US. Each bar represents a major payment instrument. The check is the oldest major non-cash instrument; it is also still the most widely-used.[25] The credit card was invented in 1949.[26] This product involves substantial network externalities: it needs a certain amount of accepting merchants and card holders to work. As a result, early schemes like Diners Club and Amex focused on the Travel and Entertainment (T&E) niche; it was relatively easy to gain penetration among travelling businessmen and the hotels and restaurants they frequented. During the first twenty years the product stayed in this niche. The advent of the open networks of Visa and MasterCard in the early 1970s laid the ground for wider usage; by 1974 the number of US credit card transactions reached the level of 1 per person per month. Since then usage has continued to grow, and it now stands at more than one credit card transaction per person per week. Next in appearance was the ATM. This instrument too is subject to network externalities. However, even a single ATM provides significant value to both consumers and banks.[27] This

[24] Most of Northern Continental Europe (Germany, Belgium, Scandinavia) went through the same payment system developments as the Netherlands.

[25] Checks were introduced in the 1920s to replace the older draft instrument. For an interesting description of this process and the factors driving it, see Prescott and Weinberg (2000). In essence they convincingly argue that widespread use of checks by businesses became feasible once a certain national integration of the US was achieved, so that the credit worthiness of the check writer in another part of the country could be relied upon and verified (a draft relies on the credit worthiness of the *bank* on which is drafted, while a check relies on the credit worthiness of the *writer* itself).

[26] Based on the history of credit cards as described by Mandell (1990) and Evans and Schmalensee (1999).

[27] See the empirical work of Hannan and McDowell (1984), Paroush and Ruthenberg (1986), Sharma (1993) and Saloner and Shepard (1995).

allowed banks to introduce them stand-alone or in small closed networks. Only later, during the early 1980s were the regional and national ATM networks established, giving the product its full 'network' value.

By the late 1980s, the US had widely adopted two network payment instruments: ATMs and credit cards. The next technology to arrive was the debit card. Like ATMs and credit cards, debit cards are subject to strong network externalities. In this case however, the network (card users and accepting merchants) was not established from scratch, but by building on existing networks. Two different debit card technologies were introduced. The banks (who owned the ATM networks) introduced PIN-based debit. This technology uses the existing ATM cards as well as the underlying ATM-networks and protocols (hence the use of PIN). Thus the banks only had to penetrate one side of the equation: the merchants. The credit card networks (Visa and MasterCard) introduced signature-based debit. This instrument uses the existing credit card networks, terminals and merchant contracts (hence the use of signature). As a result, the card companies too had to penetrate only one side of the equation: the consumers.[28] Perhaps by leveraging the existing base, debit cards overcame the chicken-and-egg problem faster than credit cards: it took debit cards about 10 years to reach the level of 10 transactions per capita (1986-1997), while it took credit cards almost 25 years to achieve that level (1950-1974). Even so, early debit card usage was highly concentrated in 'pockets': in 1993 half of all US debit card transactions were in California, and 70% of transactions were made at supermarkets and gas stations.[29]

1.3.2 Netherlands: four weddings and a funeral

The Netherlands followed a different trajectory, described in figure 1.3. A crucial difference with the US occurred early on: the development of a Postal giro-system in 1918. The banks introduced their own ACH system in the 1960s. And throughout the following 25 years the banks acted jointly (but without

[28] All merchants that accept Visa and MasterCard credit cards must also accept their debit cards under the "honor all cards rule". Following a major class-action lawsuit by Walmart, this rule has now been dropped by Visa and MasterCard as part of $3 billion settlement in April 2003 (taken from www.visa.com and www.walmart.com). These damages stem from the fact that there is an important economic difference between the two types of debit. Whereas the merchant discount for PIN debit is generally $ 0.10-0.20 per transaction (in the US), the discount for signature debit is generally the same as for credit cards, around 1-2% of the transaction amount, or $ 0.50 to 1.00 for an average transaction.

[29] See Caskey and Sellon (1994) for these figures and an extensive description of the early days of US debit cards. Interestingly, few authors on network externalities in payments point to the need to find such niches to overcome the chicken-egg problem. A notable exception is Stone (1994), who stresses the need for segmentation to get ACH technology adopted.

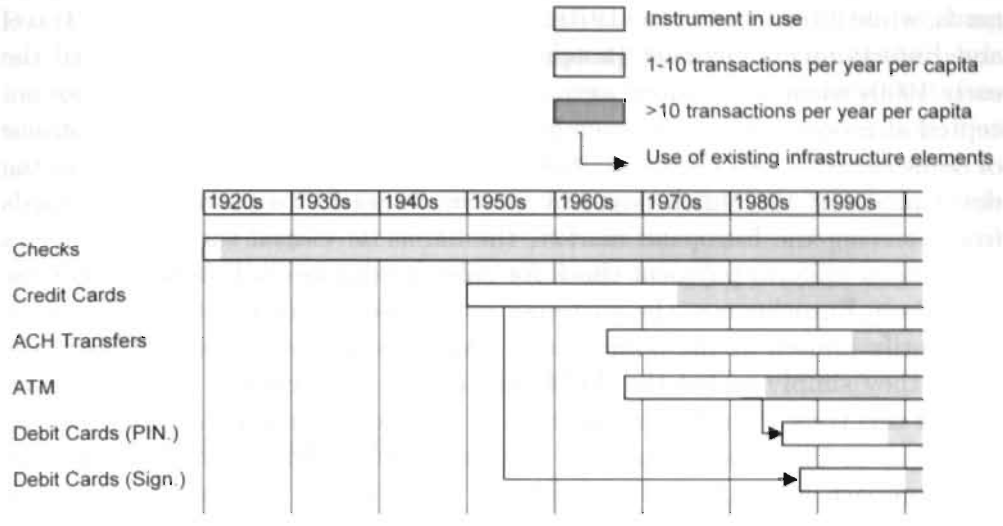

Figure 1.2 Evolution of payment instruments in the US
Source: BIS, Evans and Schmalensee (1999), Prescott and Weinberg (2000).

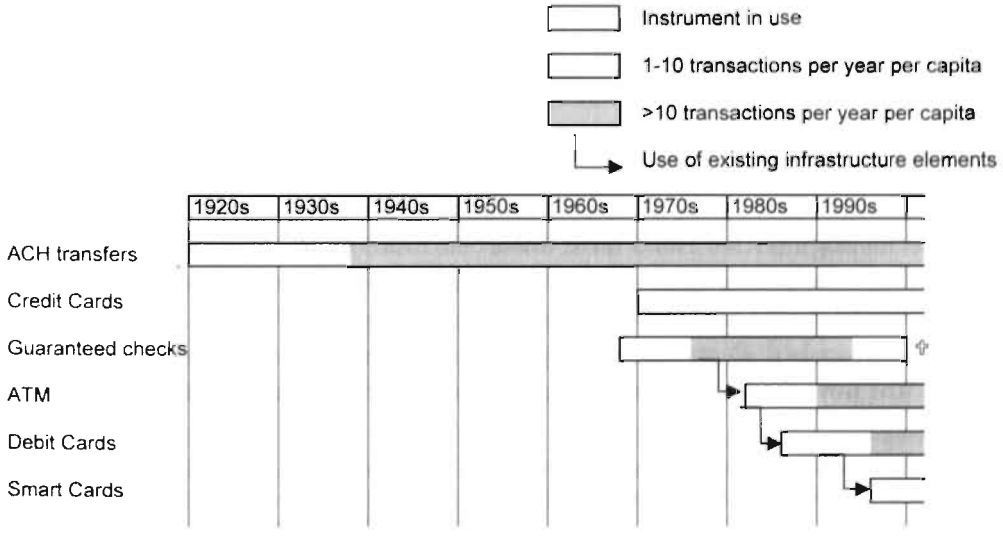

Figure 1.3 Evolution of payment instruments in the Netherlands
Source: Interpay, PCGD (1973), BIS (2003).

the Postal giro-system) in the adoption of new payment technologies. Credit cards, while introduced in the 1970s, did not gain acceptance beyond the Travel and Entertainment segment (hotels, restaurants and luxury shops) until the early 1990s when gas stations were added. Even today, credit cards are not accepted at grocery chains and merchant penetration is only half the acceptance of debit cards.[30] As Lelieveldt (2000) puts it in his English description of the development of the Dutch payment system: "in order to prevent credit cards from entering the European market, the banks developed the Eurocheque, a guaranteed uniform payment check for cross-border use in Europe".[31] For use with these Eurocheques, the customer got a 'cheque card' with his signature to identify himself at the point of sale. When banks introduced ATMs in the 1980s they simply added the ATM function to these cheque cards. They repeated this trick with the introduction of debit cards in the late 1980s, in the same way as US banks introduced PIN-based debit. The banks then repeated this approach for a third time, with the introduction of e-purse/smart card technology in the late 1990s, adding the e-purse function to both the debit card and the merchant terminal.[32]

Finally, and in contrast to the US, the Dutch payments industry saw the withdrawal of an instrument: last year, in an act of active euthanasia, the Dutch banks stopped issuing and accepting the Eurocheque.

1.3.3 Comparison of payment instrument succession in US and Netherlands

Closer analysis of two countries suggests that a major difference (the use of checks versus ACH/giro) can be traced back to events that occurred almost 100 years ago. Ever since that event, the two countries appear to follow distinct technology paths. Both have absorbed more or less the same underlying 'technology elements', e.g. by moving from cash to paper based systems: until the 1980s checks and girotransfers were largely paper based.[33] Later, both countries absorbed electronic processing and telecommunications technologies: checks are now scanned, cards have magnetic stripes and terminals relay the

[30] In 2001 Credit Cards were accepted 82,000 Dutch merchants, while Debit Cards were accepted at 165,000 locations (source: Interpay).

[31] This product (mainly driven by the German banks) differs from the American personal check because the payee bears no risk of a 'bounced check': they are always reimbursed by the issuing bank (up to a maximum of around EUR 200).

[32] To date usage has been disappointing, although from 2001 to 2002 usage doubled to 5 transactions per person per year, as parking meters and vending machines started to accept smart cards (source: Interpay).

[33] Even the early credit cards were real 'cards', being made from thick paper, not plastic.

data on-line (even for credit cards where this would not be strictly necessary). But while the underlying technological elements are the same, the resulting 'technology paths' are quite different. Compared to the US, the Netherlands has adopted and dropped more payment technologies. Convergence between the two countries is at best very slow.[34] Furthermore, there are subtle, but important, differences below the surface. For example, the US is moving to signature rather than PIN-based debit: in 2000 two-thirds of all debit cards in the US were signature- rather than PIN-based debit, almost reversing the situation of 1993 when signature debit was only 40%.[35]

1.4 Economic Relevance

Payments represent a significant part of the economy. Humphrey, Pulley, et al. (2000) estimate that the United States spends $225 billion per year just to make payments, or 3% of GDP. This (huge) figure is based on an average cost per non-cash transaction of $2.60 times 87 billion non-cash transactions per year (1996). It excludes the cost of cash, but on the other hand it still uses the old check estimate (66 billion checks per year, which was later revised downward by 17 billion)[36]. For other countries, the percentage of GDP is likely to be lower: these countries generally use cheaper instruments than checks and the number of non-cash transactions is lower than in the US. The accounting firm KPMG (1990) calculated that Dutch banks spent EUR 2.4 billion on payments in 1989, or 1% of GDP. This is probably an underestimate of total cost to Dutch society, since it excludes the cost of payments to consumers, businesses and merchants.[37] De Grauwe, Buyst, et al. (2000) do include the cost to the merchant and get a cost of 0.7% of GDP for cash alone. 0.2% of this is borne by banks, leaving 0.5% for consumers and merchants. Adding this to the figures of KPMG gets us in the range of 1.5% of GDP. Overall these estimates indicate that the cost of the payment system represents something in the range of 1.5-3% of GDP for developed countries.

---

[34] See also the results of the earlier Ben-David convergence test, keeping in mind that the Netherlands is representative for a larger group of countries.

[35] BIS (2002) and Caskey and Selon (1994).

[36] Charkavorti (2000) puts the social cost of cash at 60 billion (or 0.7% of GDP) for the US, based on a study by the US treasury. This is in line with De Grauwe, Buyst, et al. (2000). They calculate the total system cost of cash for Belgium at 0.68% of GDP.

[37] Another study (Jaarsma and van Rijt-Veltman, 2000) estimates the cost of payments to merchants at EUR 535 million in 1998. But since these costs include bank fees, they cannot be simply added to bank costs.

TABLE 1.3 Estimated cost in EUR per non-cash transaction according to various authors

Source	Cash	Check	ACH paper	ACH elec.	POS dt
De Grauwe, Buyst, et al. (2000)	0.58				0.65
Wells (1996)/Humphrey, Keppler, et al. (1997)		$2.93		1.31	
Flatraaker et.al. (1995)		1.97	0.92	0.49	0.63
Humphrey&Berger (1990)					1.75
KPMG (1990)		n/a	1.39	0.24	0.65

Costs vary significantly across the various instruments. Most of these costs are hard to measure: they represent processing costs for banks, and handling costs of merchants, businesses and consumers. None of these are readily obtainable. Several authors have tried to estimate costs of the main payment instruments. Humphrey, Willesson, et al. (2003) and van Hove (2003) give a comprehensive overview of the state of the art in estimating payment costs. Both articles highlight the pitfals and difficulties in obtaining estimates that are comparable across countries. Keeping this mind,table 1.3 gives an overview of some important estimates.[38]

Payments therefore represent a significant part of the economy and there are significant differences in costs across instruments. The impact of using different instruments can therefore be large. For example Humphrey, Pulley, et al. (2000) estimate that the US could save 1.25% of GDP if they were to move from paper checks to electronic giro.[39] Similarly, Humphrey, Kim, et al. (2001) estimate that Norway could save 0.6% of GDP by moving all of its paper-based instruments to electronic versions.

[38]The cash figure from De Grauwe, Buyst, et al. (2000) is their figure for Belgium; it includes costs to all parties: consumers, banks and merchants. The same holds for figures of Wells (1996), Humphrey, Keppler, et al. (1997) and Humphrey and Berger (1990). Figures from Flatraaker and Robinson (1995) and KPMG (1990) represent only the costs to banks. Finally the figure from Flatraaker and Robinson (1995) for an ACH paper transaction is a weighted average of their costs for several instruments including a mail giro (0.98 EUR).

[39]Stavins (1997) discusses the potential savings by moving to electronic check presentment with truncation (which would make it unnecessary to ship paper between banks), and finds them surprisingly low: if all checks were to move to this technology it would save 2.4 cents per check or $1.4 billion per year.

1.5 Questions for this thesis

Several important observations can be made in reviewing the payments landscape as it has been described in the previous sections. In the first place, we see a rapid growth of non-cash instruments, with several new instruments being adopted over the past 30 years. Second, these differences are economically relevant. Third, the use of payment instruments follows national boundaries; and while the growth of non-cash instruments is universal, there remain significant differences in the use of payment instruments across countries. Fourth, quantitative analysis reveals that any convergence across countries is slow at best, both in the overall use of non-cash instruments, and in the use of specific instruments. And finally, analysis of the successive adoption of instruments by country suggests that the arrival of new technologies did not smooth out the country differences; instead countries appear to follow different technology paths.

These observations lead to the following questions, which will be the focus of this thesis:

1. Given that most countries came from a similar background in terms of payment instruments (cash, or, going back even further, shells, salt and beads): what caused the initial differences? In particular why did some countries adopt ACH/giro-systems where others, like the US, did not?

2. Why do these differences persist, given the large economic gains that could be obtained by switching to more efficient mixes?

3. Why do countries appear to follow different technology paths in terms of the adoption of payment instruments?

4. How are these differences likely to evolve in the future?

The answers to these questions are relevant for several reasons.

- *Economic impact.* As described earlier, the choice of payment instrument has significant economic implications.

- *Implications for competitive environment.* Part of this thesis examines the potential impact of industry structure (essentially concentration) on the choice of payments infrastructure. The reverse relationship is also relevant, however. For example, credit cards can be (and are) easily issued by specialist players (MBNA, Citibank); debit cards by contrast, have to

be issued by the institution that holds the basic salary account.[40] Furthermore, giro-systems, with their standing orders and direct debits, may increase consumer-switching costs.[41] Thus, taken together, the reliance on debit cards and giro-systems (as observed in many European countries) may well increase the grip of traditional banks on their customers.

- *Likelihood of a single Euro payment infrastructure.* Last year, much of Europe switched to a single physical currency. This is not the same as one payment infrastructure. Cross-border payments are costly to banks and cumbersome to firms and consumers.

- *Convergence of European banking markets.* Different payment systems may make it more difficult for banks to acquire customers in other countries. For example, a bank may have to connect to the local payment infrastructure, develop local payment solutions, get the necessary expertise, licenses etc.

- *Insight into standardization and adoption process.* Understanding how countries adopt and standardize network technologies may increase the insight into the likely adoption of future payment technologies.

In the next part of this thesis (Part II) I will turn to question 1 and 2: what caused initial differences and why do they persist? Chapter 2 reviews the existing literature on payments and network economics, because this is found to be the most likely explanation of country differences. Chapter 3 introduces a model to analyze the effect on the adoption and compatibility decision by firms of: (a) industry structure, and (b) the role of countries, assuming that standards are unsponsored. Chapter 4 does the same for sponsored standards. This model is applied to two cases from the payments industry: the introduction of ACH/Giro in the Netherlands (chapter 5) and the harmonization of transfer payments in the Eurozone (chapter 6). Part III looks at the other two questions: why do countries follow different technology paths and what are implications for future adoption? Chapter 7 analyzes the existing literature on innovation and technology succession. Chapter 8 describes a model for succession of network technologies and uses it to analyze technology paths across different countries. Chapter 9 uses the results to analyze the case of Internet payments. Part IV summarizes the conclusions and speculates about further applications.

---

[40] As pointed out by i.a. Evans and Schmalensee, (1999).

[41] For example, the Dutch competition authority (NMa) recently argued that banks should offer their customers the ability to take their account number with them to another bank (NEI, 2000).

2
Theory of payment instruments and networks

Section 2.1 reviews the theory on payment instruments, looking for answers to the questions regarding differences in the use of payment instruments between countries. Several authors acknowledge that traditional micro-economic mechanisms (price, consumer characteristics) fail to adequately explain the country-differences. Instead they suggest network externalities could play a role. Section 2.2 therefore examines the literature on network externalities, lock-in and path dependence, mostly developed in the late 1980s and early 1990s. Section 2.3 then looks at the applications of this theory to payments and finds substantial evidence that many payment instruments do indeed exhibit network externalities. The last section of this chapter summarizes the conclusions from this literature overview.

2.1 Payment instruments

Payments and money are closely related, and money is one of the crucial ingredients of economics. It will therefore come as no surprise that the body of literature on payments is large indeed. In his (far from exhaustive) review of payment systems literature, Khiaonarong (1997) reviews over 350 sources, including many newspaper and magazine clippings. Limiting themselves to economic literature, Hancock and Humphrey (1998) cite 130 sources in their thorough review of the literature on "payment transactions, instruments and systems". They find the literature focuses on three topics: (1) the trade-off between cash and non-cash instruments; (2) the use of different non-cash instruments and the implications for money, bank deposits and monetary policy; and (3) the risk of settlement failure in large value payment transactions (systemic risk).

The topic of this thesis is not money or the balances people hold. The main concern is the choice of payment instrument. For example: why do people use credit cards instead of writing a check or paying with cash? And why does the use of instruments differ across countries? I am therefore mainly interested in the first topic of Hancock and Humphrey (the trade-off between cash and non-cash instruments) and part of the second topic (the use of non-cash in-

struments). While interesting, the monetary consequences of instrument choice are not my primary concern. The same holds for the risk of settlement failure in large value payment transactions, since my focus is on retail instruments.

2.1.1 Cash versus alternatives

Baumol (1952) analyzed why people hold currency (notes and coins) instead of other forms of money, notably bank deposits.[1] He used an inventory model, where a consumer ('representative agent') makes a trade-off between the cost of a trip to the bank to get cash and the foregone interest revenue on cash. The cash inventory follows a sawtooth model, as balances are replenished by a cash withdrawal, then gradually spent on goods an services. The optimal cash balance m is given by $m = \sqrt{\frac{cT}{2r}}$, where c is the cost of replenishment (cost of a trip to the bank), T is the total expenditure during a period, and r is the interest rate on any balances with the bank. It easily follows that the income elasticity of cash balances should then be $\frac{1}{2}$, but most authors find smaller elasticities.[2]

Building on Baumol's work, Prescott (1987) studies the trade-off between two means of payment: cash and drafts (checks) written against an interest bearing account. There is a fixed cost of writing a check, while cash has an opportunity interest cost. In equilibrium cash will be used for small purchases and checks for larger ones. Prescott uses this model to compare the effect of a country's welfare on equilibrium. He finds that in a rich country (with a higher marginal product of capital and greater output) the use of checks of is far greater than in a poor country. Since checks are socially costly (they require resources to produce) while cash is not, the optimal nominal interest rate is zero; in that case checks won't be used, since consumers do not get interest on

[1] Keynes distinguishes three motives for holding money balances (the transaction, precautionary and speculative motives). These are however motives for holding *money*, which includes both currency (notes and coins in the hands of the public) and bank deposits.

[2] This assumes the expenditures are proportional to income. The income elasticity is then derived as follows:

$$\varepsilon = \frac{\partial m}{\partial T}\frac{T}{m} = \left(\frac{1}{2}\sqrt{\frac{c}{2r}}\,(T)^{-\frac{1}{2}}\right)\frac{T}{m} = \frac{1}{2}\left(\sqrt{\frac{cT}{2r}}\,(T)^{-1}\right)\frac{T}{m} = \frac{1}{2}.$$

For a comprehensive survey of this topic, see Boeschoten (1992). Using data from a Dutch consumer survey, he finds elasticities in the range of 0.15-0.35. He quotes several other sources that all report elasticities that are lower than $\frac{1}{2}$. He concludes that the relationship is weaker than Baumol's model predicts.

their checking account but still face the fixed cost per check written.[3] If the nominal interest rate is positive, checks will be overused (from a social welfare perspective), and cash will be underused.

My concern with the models of Baumol and Prescott is that most banks do not give interest on checking accounts. Foregone interest may explain *money* balances (cash plus bank deposits) but it is an unlikely explanation for the size of *cash* balances. An inventory model may still apply, but the cost of holding cash has probably more to do with the risk of loss or theft than with interest. Even if banks would give interest on checking accounts, the average US consumer household holds about $100 in cash, representing at most $5 per year in foregone interest; this is almost negligible compared to the $50-150 in fees that US banks charge to maintain a checking account.[4]

While Prescott's model leads to cash being under-used, ten Raa and Shestalova (2003) reach the opposite conclusion. They use detailed data on the cost of cash and debit card payments from 215 Dutch merchants, including operational costs (telecommunication, POS terminals, transportation) and foregone interest. Regressing the cost per merchant on the number and value of transactions, the authors find:

$$\begin{aligned} c_{cash} &= 0.019 + 0.25\%\ v \\ c_{debit} &= 0.060 + 0.11\%\ v. \end{aligned} \tag{2.1}$$

Here c_{cash} and c_{debit} are costs in euro per transaction of cash and debit cards respectively, while v is the value of the transaction, also in euro. It easily follows that the break-even point for the merchant is about EUR 30: for transactions above this level the merchant prefers payment by debit card, while cash is the cheaper means for smaller payments. The authors then add the costs to banks to get the social costs of these instruments.[5] This lowers the break-even point to EUR 13. As a result, merchants will prefer cash for transactions between EUR 13 and 30, where debit cards would be cheaper from a social point of

[3] As Prescott points out this is in line with an assertion of Milton Friedman that, in an otherwise stable economy, the optimal growth rate of currency supply (and thus inflation) is minus the real interest rate.

[4] Figures on cash holdings taken from Porter and Judson (1996), table 1 and Boeschoten (1998). The figures on the cost of a checking account come from PIRG (1999); their analysis shows that it costs an average of $212 per year to maintain a checking account at a US bank and an average of $112 at credit unions. These costs include about $30-50 in foregone interest on the account balance, but the rest consists of fees, charges and commissions.

[5] Ten Raa and Shestalova argue that since the banks do not charge consumers for either instrument (ATM withdrawals and debit cards are free in the Netherlands), the cost to the consumer is zero. For the banks' cost of cash they use the Norwegian data from Flatraaker and Robinson (1995), while debit cards are supposed to be cost neutral for banks (the merchant fee covers the cost to banks).

view. Since the authors assume the merchant has an important influence over the choice of payment instrument, ten Raa and Shestalova conclude that cash is overused. While plausible, I am not convinced this reasoning can explain country differences. In France for example, the merchant pays a banking fee per debit card transaction equal to about 0.8% of the transaction value, compared to a flat fee of 0.04 euro in the Netherlands. This would change expression (2.1) to:[6]

$$c_{debit} = 0.02 + 0.91\%\ v,$$

implying cash is always cheaper for a French merchant. Since both instruments are free to the French consumer, one would expect debit card usage to be much lower than in the Netherlands. In reality it is about the same.[7]

While cash may be optimal (from a social welfare perspective) for small purchases, in practice it is often used for large payments. As Hancock and Humphrey (1998) state, the advantages of cash include that it represents final payment, it is immediately reusable, and it is divisible; to which I would add the advantage that it is anonymous. Thus, in spite of the obvious disadvantages of cash (risk of theft and loss), one expects it to be used for specific occasions like cattle markets (where immediate reusability and finality is crucial) and illegal purposes such as drug traffic and tax evasion. This pattern is confirmed by a Dutch survey described in Boeschoten and Fase (1992), who find that the majority of large bank notes (denomination of NLG 1000 or about EUR 500) is either hoarded or used in agriculture, drug trade, car trading and gambling. The authors estimate that a third to half of all Dutch currency is hoarded, a fact which they largely attribute to tax evasion.[8]

2.1.2 Will cash disappear?

Quite a few authors analyze the effect of non-cash instruments on the demand for currency and money. A drastic reduction would mean a loss of seigniorage revenue to governments (cash represents an interest-free loan to the government), while a structural reduction in the demand for money could reduce the effectiveness of monetary policy. For example White (1976) and Daniels and

---

[6] Compared to expression (2.1) the fixed cost is lowered by EUR 0.04, while the variable cost increases with 0.8%.

[7] I can think of two explanations for this: (1) the merchant has very little influence over the choice of payment instrument used by his customer (quite plausible); and (2) cash is more expensive to a French merchant than to a Dutch merchant (could be, but I have no evidence of this).

[8] The authors calculate that at a tax rate of 50% and a return of 4% on savings deposits, it is more profitable to hold cash (while evading taxes), as long as the holding period is less than 35 years.

Murphy (1994) find evidence that households that are heavy users of credit cards hold lower currency balances. Duca and Whitesell (1995) find the same relation for credit cards and demand deposits, while Avery, Elliehausen, et al. (1986) find that use of ATMs reduces currency holdings.[9] Boeschoten and Hebbink (1996) calculate that the replacement of all transaction balances by electronic money would reduce seigniorage revenues by up to 0.7% of GDP. Costa and De Grauwe (2001) are already speculating about the consequences of a true cashless society: what would be the unit of account, how can monetary policy be conducted?

This all seems a bit premature because a large part of currency is either hoarded or used for activities where cash is not easily replaced by non-cash. Boeschoten (1998), using survey data from Dutch households, finds that increased usage of POS/debit could reduce transaction balances by up to 40%, but argues that the eventual impact on money demand will be much smaller, due to the large stock of currency held as hoards.

As Humphrey, Kaloudis, et al. (2000) point out this implies that an increasing share of cash is used in the 'grey economy'. Using Norwegian data they forecast the use of cash and conclude: "Overall, our results suggest that cash used in legal activities may become so small in Norway that government seigniorage revenues are likely to be due almost solely to providing the means of payment for illegal activities". This result is confirmed by Drehmann, Goodhart, et al. (2001), who conclude that "indeed, the legalization of drugs could make a much bigger dent in the demand for currency than competition from e-money".[10]

2.1.3 Choice between non-cash instruments

Thus the trade-off between cash and non-cash instruments is fairly well understood, and seems to be largely driven by the size and 'context' of the transaction. The choice between various non-cash instruments is more difficult. A comprehensive theoretical model is offered by Santomero and Seater (1996). They follow an approach similar to Prescott (1987, described earlier in section 2.1.1), but the model of Santomero and Seater can cope with any number of payment instruments, where Prescott's model included only cash and checks. Consumers can either keep money in a savings account at some interest r_s, or in several other accounts, each enabling the use of a specific payment instrument.

---

[9] The survey results of Avery, Elliehausen, et al. (1986) also show that 14% of US households use cash as their *only* means of payment. This includes the bankrupt and many illegal aliens.

[10] Drehman, Goodhart, et al. (2001), p. 217.

All these accounts yield an interest lower than the savings account. For example, cash could be considered an account with zero interest, a checking account would offer some intermediate rate, and (with a little stretching) one could argue that a credit card account yields a negative rate on a negative balance. These payment instruments can be used to purchase goods (make shopping trips). Each of these goods comes in lumps, and the consumer faces several decisions: how much money to keep in the savings and payment accounts, which instrument to use for which good, and how much to purchase of each good (the goods can be stored till consumption, yielding some return). Using a representative agent approach the authors derive equilibrium. They find that the instrument with the higher interest rate will be used for the larger expenditures. Thus even consumers with the same income, but different purchase patterns, may use a different mix of payment instruments. The relationship with income is highly dependent on the parameters of the model. For some values, higher income households use more instruments, but for other parameter values the relationship is ambiguous. The authors conclude that payment instruments behave like cars (or any other consumer good): there is a demand for different varieties, and they finish their article with a warning: "The choice of money or monies to be used for transactions purposes is a complex decision. It does not lend itself to simple extrapolation from consumer surveys, and, in fact, may result in substantially different outcomes than had been presumed. The innovators would do well to proceed slowly."[11]

It is hard to argue with these general conclusions. Nevertheless I find the Santomero and Seater model problematic in two ways. First, the level of abstraction is very high: it is not obvious to translate their model into actual payment instruments. More importantly, equilibrium in their model depends only on income and purchase patterns. It cannot explain differences across countries, unless we assume that purchase patterns and incomes differ quite dramatically among countries of the developed world.

Shy and Tarkka (1998) analyze the choice between three concrete instruments: debit cards, currency (cash) and e-purses (on a smart card). Each of these offers different transaction costs. For example, a debit card payment requires the verification of the account balance by phone, while an e-purse payment does not. Currency can be lost or stolen, while there is a time loss to the merchant because coins need to be counted, and change given; and e-purses bear the risk that the card may malfunction ('magnetic failure') in which case the balance is lost. Banks set fees to distribute these costs over consumers, merchants and themselves. The authors assume e-purses are attractive to banks

[11] Santomero and Seater (1996), p. 959.

(because they get the interest on the balance) so banks will set consumer fees low enough to get them used. In equilibrium, debit cards will be used for large transactions, e-purses for small transactions, and currency for the intermediate range. The authors show that these ranges do not depend on whether the supply of e-purses is competitive or monopolistic. However, there is market failure because currency is overused in equilibrium.[12]

The model of Shy and Tarkka certainly addresses my first objection to Santomero and Seater: it relates easily to real observable payment instruments. However, it does not address my second objection. In the model of Shy and Tarkka all countries should end up with the same instrument mix. In particular, the authors assume that e-purses are welfare improving (if used for small transactions) and will be introduced. There is no room for the type of market failure where an instrument is not introduced even if it improves social welfare.

Several other authors offer theoretical and empirical explanations for the use of specific instruments. In general, two types of explanations are offered (1) characteristics of the instrument, and (2) characteristics of the user.

Characteristics of the instrument

Humphrey, Kim, et al. (2001) analyze data on Norwegian payments, covering 1989-1995. The use of payment instruments shifted dramatically during that period. The use of checks fell by 70%, while the use of POS debit rose six-fold. The authors argue that pricing played a major role. The consumer price of an ATM transaction tripled from EUR 0.18 in 1989 to 0.51 in 1995, the price of a check doubled from EUR 0.55 to 1.13, while the price of a POS debit transaction rose by only a third, from 19 to 26 eurocents. Regressing transaction shares on these prices, the authors find own price elasticities of 0.96, 0.75 and 0.87 for ATM cash withdrawals, checks and POS debit respectively. However, other European countries made similar shifts without the explicit pricing of the Norwegians. Moreover, other authors find much lower price elasticities: using cross-country data, Humphrey, Pulley and Vesala (1996) find own price elasticities of 0.09-0.26; Murphy (1991) finds that per item charges on checks reduce

[12] This is because the merchant bears a large part of the cost of currency (he needs to count and store the cash), while banks make e-purses attractive to consumers (since banks get the interest on the balance). And so the debate on the over- or underusage of cash continues. Lacker (1996) claims it is underused: the foregone interest is not a real cost but a value transfer to the government, yet the private sector engages in socially costly activities like e-purses to avoid the interest loss. Prescott's (1987) theoretical model leads to the same conclusion, through the very mechanism that Lacker describes. Shy and Tarkka (1998) and ten Raa and Shestalova (2003) both conclude that currency is overused, but for different reasons: the former because the *consumer* prefers currency, the latter because the *merchant* prefers currency.

usage by only 10%; and several authors have been puzzled by the willingness of consumers to pay high interest rates on credit card debt.[13]

Price distortions are also claimed to play a role in the continued usage of checks in the US: the writer of a check gets the benefits (primarily float), while the recipient bears the cost. This argument is elaborated in Humphrey and Berger (1990). But the role of price is again disputed: Wells (1996) argues that the average float has declined significantly, from $ 1.04 per check in 1987 to $ 0.09 in 1993, and is anyway insignificant for the average consumer check.[14] And indeed, with US banks charging an average penalty fee of $20 per bounced check (PIRG, 1999) it seems hard to believe that US consumers write checks just because of the float advantage.[15]

The above models all base instrument choice purely on cost and opportunity returns, which seems an incomplete set of attributes. Hirschman (1982) conducts market research where over 1000 consumers assign 11 attributes to 5 payment instruments. These includes things like documentation of the transaction, acceptance, etc. She concludes that these attributes are linked to instruments by consumers across all demographic segments. Consumers indicate consistent preferences for using specific instruments depending on the type of purchase or payment. For example, cash is seen as having a low transaction time and universal acceptance, and is used for small day-to-day purchases. Credit cards are seen as secure, prestigious while they inhibited the ability to control spending, and are therefore used more sparingly than cash (Hirschman's conclusions aren't really shocking).

Characteristics of the user

There is ample evidence that demographics play an important role in instrument choice. In general the young and affluent are more inclined to use electronic instruments. Mantel and McHugh (2001) find debit card usage to be higher among the young and affluent, but not correlated with other 'new product adoption factors' such as mobile phone usage and Internet shopping; they

[13] Ausubel (1991) finds that interest rates charged on credit card debt are not significantly related to the cost of funds of credit card companies: over the years 1982-1989 they remain stable at just under 20%, while the cost of funds fell from 15% to 10%. Since there are 4000 firms in the US credit card market, Ausubel rules out firm cooperation as a cause. Using more recent data Gross and Souleles (2002) do find a significant price sensitivity, with an elasticity of about 1.3; they are puzzled however by the fact that many people simultaneously borrow and hold on to low yielding assets.

[14] Real dollar terms: figures are corrected for inflation. The decline is due to lower interest rates and quicker processing.

[15] The PIRG (1999) report mentioned earlier puts the average bounced check penalty at $22 for banks and $17 for credit unions.

also find debit usage higher among inhabitants of large markets (more than 500,000 people), which, as the authors suggest, may be due to network externalities. For credit cards that same study finds no relation with market size (perhaps because the networks are mature), but a relationship with race: non-caucasians are significantly more likely to borrow on their card. As far as POS/debit is concerned these results are in line with the findings of Boeschoten (1998) who finds that the usage of POS debit is significantly higher for the young, the affluent, and the urbanized. Using perhaps the largest survey, the 1998 Survey of Consumer Finances (SCF) among 4,305 families, Stavins (2001) finds the same overall pattern. In addition, for all electronic instruments she finds a significant influence of the habits of other consumers nearby. She interprets this as a possible indication of demand related network effects.

2.1.4 Explaining country differences

None of the literature reviewed in the previous section focuses on country differences. They all either derive some equilibrium, based on income, expenditures, interest rates etc., or relate usage to characteristics of the instrument and user. However, since none of these factors is obviously different across countries, these models cannot satisfactorily explain country differences.

The most comprehensive study of such cross-country differences in the use of payment instruments is made by Humphrey, Pulley, et al. (1996), using data on five instruments (checks, credit cards, debit cards, paper giro and electronic giro) over 7 years (1987-1993) and across the G-10 countries. The authors regress annual per capita transactions for each instrument on explanatory variables like the price of the various instruments, per capita income, number of ATMs and debit card terminals, currency holdings per person, crime rate and bank-sector concentration. While they find significant relationships with variables like number of ATMs and debit card terminals, they acknowledge that these variables are largely endogenous. On the other hand the explanatory value of truly exogenous variables like crime rate and income is quite modest. While the authors show that the availability of places to use an instrument (e.g. number of EFTPOS terminals) plays an important role, they stop short of suggesting network externalities and/or lock-in as an explanation. Overall, the authors conclude that while there is a common trend from cash to non-cash and from paper to electronic transactions, there are large and persisting differences that appear to be caused by country 'idiosyncrasies'.

The most notable 'idiosyncrasy' is the continued and intense use of checks by Americans. In their 2000 article, Humphrey, Pulley, et al. analyze this "American love affair with checks". In the best romantic tradition, this love affair appears to be both irrational (i.e. socially costly), and persistent. Several ex-

planations are offered, such as perverse pricing (not only are checks free, but a consumer gets the float on a check, since it takes a few days before his account is debited), convenience for consumers, and sunk investments in processing checks. None of these appear decisive. For example, it is not clear why Americans find it more convenient to use checks compared to consumers in other countries. As for the sunk investments, two findings suggest that this effect may be smaller than some authors assume. First, scale economies in check processing are exhausted at fairly low volumes. Second, over the last decade, studies have not found evidence of significant technological change in check processing.[16] Both findings suggest there has been little economic need to invest in large-scale infrastructure for processing checks.

Perhaps the most convincing explanation offered by the authors is the fragmentation of US banks, and strong anti-trust regulation. Any individual bank has little incentive to move to electronic transfers, while anti-trust regulation complicates coordination among banks. This is in contrast with Europe: "Examples of such cooperation include ... banks' joint ventures to develop and manage giro ... networks in a number of European countries".[17]

The issue of continued US check usage is revisited by Chakravorti and McHugh (2002) who stress that most consumer benefits are already offered by checks, so the consumer incentive to adopt alternatives like ACH are limited. In addition they mention the chicken-and-egg-problem of ACH adoption.

To summarize, factors like price, payment occasion and demographics may explain the choice of non-cash instrument, but they cannot adequately explain cross-country differences. Instead the payments literature indicates that the most promising explanations for country differences are to be found in the network externalities of many payment instruments. I therefore now turn to the literature on this topic.

2.2 Network externalities

'Network effects' have received wide attention in the late 1980's.[18] For a comprehensive overview, see David and Greenstein (1990) and more recently Shy (2001). From the start it was recognized that network externalities can give rise to multiple equilibria. Therefore much of the literature tries to answer three

[16] See Bauer and Ferrier (1996) for both findings.

[17] Quote from Humphrey, Pulley, et al. (2000).

[18] Rohlfs (1974) already describes the main characteristics, like multiple equilibria, using a model very much like one used later by Economides and Himmelberg (1995). For some reason his contribution went largely unnoticed.

questions: (1) under what circumstances do network effects lead to multiple equilibria? (2) are some of these equilibria socially suboptimal? and (3) what can be done to prevent or get away from these suboptimal equilibria?

Following Arthur (1989), I distinguish unsponsored and sponsored standards. The former are available to all: any firm can join a standard. The latter are owned by a firm or group of firms, allowing these firms to use the standard as a competitive weapon.[19] I first review the literature on these type of standards, looking specifically for models that I can apply in my setting: the adoption of payment technologies. I then review the literature that analyzes the adoption of standards by spatially separated (autarkic) user groups, because the topic of this thesis is the adoption and use of different payment standards across users in various countries.

2.2.1 Unsponsored standards

The fax is perhaps the best known unsponsored standard that enabled a rapid 'exponential' adoption. Economides and Himmelberg (1995) analyze the US fax market and find strong network effects. The fax is an example of a direct network effect, where the value of joining a network depends directly on the number of other consumers who join by adopting the same technology. However many payment networks display indirect network effects: for example the value of a credit card to its holder depends on the number of merchants that accept it, not on the number of other credit card holders; but since more card holders lead to more accepting merchants, an increase in the number of card holders indirectly affects the network value to an individual card holder. Church, Gandal, et al. (2002) prove that such indirect network effects have the same effects as direct network effects.

Farrell and Saloner (1985) analyze the adoption by N firms of an unsponsored standard Y. During each of n periods one specific firm can switch from an existing standard X to a new standard Y (firm j can switch in period j). Let $B_j(S, Y)$ denote the benefits to firm j of switching to Y if all firms in a subset S switch ($S \subseteq N$). Assume $B_j(S, X) \leq B_j(S\prime, X)$ and $B_j(S, Y) \leq B_j(S\prime, Y)$ for all $j \in S \subseteq S\prime$: both standards are subject to positive network externalities. Assume also that $B_j(N, X) < B_j(N, Y)$ for all j: if everybody switches to the new standard, all players are better off than if they all stick with the old stan-

[19] There is an analogy here with other increasing returns such as Marshallian externalities and learning effects. The former cannot be claimed or captured by a single firm, and behave like unsponsored standards. The latter generally accrue to the firm itself, in the which case they behave like sponsored standards, where the increasing returns benefit the sponsor. However, learning can also accrue to all firms (learning 'spillovers') and behave more like unsponsored standards.

dard. Farrell and Saloner show that under complete information and identical preferences standardization benefits cannot 'trap' an industry into an obsolete or inferior standard when there is a better alternative. The essence of their proof, which uses backward induction, will be discussed in section 3.2.1. When they drop the sequential timing and instead make timing endogenous Farrell and Saloner get an even stronger result: there is now 'excess momentum', i.e. a bias towards switching.[20]

The conditions of complete information and identical preferences are crucial for these results. When Farrell and Saloner drop this condition, excess inertia can occur: "the switch will not be made, although it would have been made in a world of complete information, and although both firms would then be better off".[21] While this lock-in is in essence a coordination problem, communication does not necessarily improve things: "communication ... eliminates excess inertia where the preferences of firms coincide, it increases inertia where the preferences differ".[22]

David and Greenstein (1990) review the network literature and find general consensus that unsponsored standards can indeed lead to multiple equilibria and 'lock-in'. Perhaps the simplest model is that of Arthur (1989), where heterogeneous consumers arrive on the market and adopt one of two standards, depending on their own preference and the number of other users of each standard. He shows how the arrival order of the earliest adopters can lock the system into either standard, even if that standard is socially inferior. This phenomenon is known as path dependence: the system has multiple equilibria (two in this case, being the full adoption of either standard) and which one is reached depends on historical events. In Arthur's model there is incomplete information (the arrival order is random) while preferences do not coincide (there are 2 'types' of consumers). Therefore Arthur's results do not contradict those of Farrell and Saloner (1985).

An example of the practical occurrence of such suboptimal outcomes is given by David (1985). He argues that the world is stuck with an inferior typewriter keyboard lay-out (QWERTY) due to early technical considerations (frequently used letters were put apart to prevent jamming of the machine). Along similar lines, Cowan (1990) finds that the early adoption of nuclear reactor technology in submarines led to the use of the light-water cooling technique even for land-

[20] This is because there is value in early switching since it commits a firm. Hence firms that really want Y can force others to jump on the bandwagon, even if these others would prefer X if everybody else would also stick with X.

[21] Farrell and Saloner (1985), p. 78. The proof uses a 2-firm setting, but the authors show how to extend the results to the case of n firms.

[22] Op. cit., p. 81.

based nuclear power plants, while this technology is arguably inferior for larger land-based reactors.

Even if interests are aligned and there is complete information (players know each other's preferences), uncertainty with regard to the technology and its benefits can still spoil things. Firms trying to coordinate and policy makers face the problem that it is often not known ex-ante whether a standard is superior. This problem is analyzed by Cowan (1991) using a 'two-armed bandit model'. A choice must be made between two technologies of unknown merit, and each trial with one of the two technologies leads to an improvement of the selected technology, as well as better know ledge of its merits. A central policy maker, following a strategy (i.e. set of choices of technologies for each subsequent trial) that is socially optimal *ex-ante*, will not necessarily end up with the technology that is optimal *ex-post*. Choi (1996) points to a trade-off in the timing of the 'fixing' of a standard: too early, and better but later alternatives may be missed, too late and firms may already be lock-in to their own standards.

While most economists accept the existence of multiple equilibria, their economic relevance is debated. For practical cases it is often difficult or impossible to determine just how suboptimal the actual outcome is. Part of this is due to the difficulty of determining the relative merits of technologies even ex-post: how good would steam engines be today, if they would have been fully developed? Would electrical cars be superior if battery technology would have received the same level of resources as gasoline technology?[23] See for example the critique of Liebowitz and Margolis (1990) on the QWERTY example. Curiously, only limited attention has been given to an easy supply of counterfactuals: the different outcomes across countries. For some reason the 'macro-economists' (analyzing the role of technology in explaining differences in per capita income) have focused on this more than the micro-economists and 'network people': But few macro-economists really get into network effects, path dependence and lock-in. An exception is Krugman (1994) who shows that local 'spillovers' can lead to geographical specialization (I will cover his model and a few related others in section 2.2.3).

Overall, several authors show how unsponsored standards can lead to lock-in into an inferior standard, especially if coordination is difficult, and if interests are not aligned. If payment instruments do indeed exhibit network externalities, and if standards are unsponsored, the model of Farrell and Saloner (1985) may offer a good starting point to analyze the adoption of payment instruments by banks and/or consumers in a country.

---

[23] This case is described by Cowan and Hultén (1996).

2.2.2 *Sponsored standards*

Katz and Shapiro (1986) show how strategic pricing by duopolists helps bring forward the benefits of a superior technology, because its sponsor lowers the price to increase the number of users, and hence the value of the standard. A sponsor internalizes the externality, potentially taking away the cause of lock-in. However, as the authors show, this does not guarantee the adoption of the superior technology. Farrell and Saloner (1986) show how an incumbent firm can use its installed base to keep a newcomer with a superior technology out of the market.

In the case of sponsored standards, the compatibility decision may become endogenous: firms may or may not decide to link make their products interoperable (for example banks can decide to link their ATM networks). Katz and Shapiro (1985) show how a firm may let rivals into its network, trading-off the higher value of the network (due to its increased size) against the sharing of the profits with their rival. Again multiple equilibria are possible, with universal compatibility being the socially superior but certainly not the only possible outcome. There are several empirical examples where firms, or coalitions of firms, use standards as a competitive weapon. Sponsored standards have been empirically analyzed in areas like DVD's (Dranove and Gandal, 1999), spreadsheets (Gandal, 1994) and VCRs (Park, 2000), and in none of these cases did leading firms grant compatibility to all others.[24]

Coordination becomes more difficult with sponsored standards, since the stakes, and hence divergence of interests between firms, are larger. Besen and Farrell (1994) formalize the two-firm situation with normalized pay-off matrices, and distinguish three types of games: (1) 'Tweedledum and Tweedledee': both firms prefer to fight with incompatible standards, (2) 'battle of the sexes': each wants his own standard but prefers compatibility over a standards war, and (3) 'pesky little brother': one firm prefers incompatibility while the other wants compatibility. Only the second game has compatibility as an equilibrium outcome.[25] For this second game (battle of the sexes), Farrell and Saloner (1988) examine three mechanisms for achieving coordination: committees, markets and hybrids. Committees are slower but yield better outcomes than markets, which rely on unilateral action by a player after which the rest jumps on the bandwagon. The third mechanism combines both communication

[24] This is not to say that firms always compete on standards. For example Philips widely licensed its audio cassette and audio-CD formats.

[25] In fact, it has 2 equilibrium outcomes. However, coordinated mixed strategies (the classic game theory solution to the battle of the sexes) do not apply here, since standard setting is generally a one-shot deal. Instead, Besen and Farrell propose a variety of bargaining and compromise solutions.

and unilateral preemptive action. In a 2-firm setting Farrell and Saloner show how these unilateral actions can actually improve the committee system; it captures the best of both worlds.

Competition between sponsored standards has perhaps been most extensively modelled using some form of differentiation. The relative size of differentiation compared to the network externality is crucial. Generally, differentiation may allow different networks to coexist, each catering to a group of consumers with different preferences. However, if the network externality is large compared to the differentiation, only one network will remain; this point is already demonstrated by Arthur (1989). DePalma and Leruth (1993) explore the precise relationship for linear network benefits. A fixed population of consumers joins one of two networks. The consumers are differentiated with parameter μ.[26] The sponsor of each network faces a demand that is dependent on (1) the difference in price; and (2) the difference in the externality of both networks. If a single consumer switches from network X to Y, the difference in the relative attractiveness of both networks will change in X's favor, leading to an increase in demand for network X. If this increase in demand is one customer or more, the process 'feeds on itself' until all consumers have joined X. In that case any equilibrium with two networks is unstable. Both networks can coexist, however, if the increase in demand is less than one customer.

Economides and Flyer (1998) analyze compatibility coalitions among 'Cournot oligopolists'. N firms produce a good with benefits $k+n$, where n is the number of users sharing the same standard and k are the standalone benefits. Firms can decide to use the same standard, which raises the value of their goods. Firms play a 2-stage game: in the first stage they form coalitions, with all firms in a coalition using the same standard. In the second stage they compete à la Cournot. Economides and Flyer analyze two regimes: (1) unsponsored standards, leading to what they call "uncoordinated equilibria", where each firm can freely join and leave coalitions; and (2) sponsored standards leading to what they call "consensual equilibria", where a firm can only join a coalition if all others agree. The authors find that the outcomes depend crucially on k, the relative strength of the network externality.[27] If it is high, the net-

[26] This corresponds to the parameter μ in the Multinomial Logit and parameter t (unit transportation cost) in the address models.

[27] While the Economides-Flyer model has a structure that does not allow for direct translation into the 'DePalma-Leruth' condition, a similar mechanism is at work. If the parameter k (which in the model of Economides-Flyer is inversely related to the network strength) is smaller than 1.1, (for the 2-firm case) the network starts 'feeding on itself' ; an increase in market share has always two effects on the hedonic price: a positive effect through the network externality and a negative effect through transportation cost (or increasing disutility of a product that is different from a consumer's ideal product). Beyond a critical value, the

work externality is weak and full compatibility is the only non-cooperative equilibrium, while there are multiple consensual equilibria (of which full compatibility is one). If k is low, the network externality is strong and there is no non-cooperative equilibrium, while all the consensual equilibria involve two or more incompatible standards.

Jonard and Schenk (1999) use the MNL to analyze the compatibility decision by differentiated firms. Firms face a trade-off: compatibility increases the size of the network externality and softens competition (which increase profits), but it also reduces differentiation, which reduces prices and profits. The outcome will depend on the relative size of both effects, but socially sub-optimal outcomes occur, where firms choose incompatibility.[28]

A similar result is obtained by Shy (2001 and 2002), who uses switching costs as a differentiating mechanism. Since these switching cost are not influenced by the network externality (unlike the model of Jonard and Schenk, 1999), he finds that network owners will always prefer compatibility, because it softens competition, while consumers always prefer incompatibility.[29] Matutes and Regibeau (1988 and 1992) use an 'address' approach, where consumers differ in their preference for two versions (each supplied by a different firm) of two components of a system, e.g., amplifiers and speakers in a stereo-system. The supplying firms may or may not make these components interoperable. The authors show that for a wide range of parameters, firms will choose to produce compatible components but will also offer discounts to bundle their products.

Matutes and Padilla (1994) give an application of such address-differentiation. They consider the decision by banks to make their ATMs compatible. In their model, compatibility will enhance the value of bank services to consumers; because it also decreases differentiation between banks, it leads to increased price competition for deposits. Using an address model for 3 banks located on a Salop circle the authors find that full compatibility between all three banks is never an equilibrium. They find multiple equilibria where either all three banks choose incompatibility or two banks establish compatibility at the exclusion of the third (see 4.5.2 for a discussion of this result).

---

positive (network) effect is bigger than the negative (transportation) effect, in which case there are no more internal equilibria.

[28] This result is not trivial, since Jonard and Schenk assume downward sloping demand. Hence incompatibility will reduce the network externality, which in itself always *reduces* social welfare, but it also reduces prices, which may *increase* social welfare.

[29] For this result see Shy (2001), p. 31. He obtains almost the opposite result when he applies his model to ATMs (op. cit. p. 201), but that is due to the fact that he treats the number of ATMs as given instead of relating them to market share. If this assumption is changed, his model again shows that banks prefer compatibility.

In summary, almost all models of sponsored standards focus on the compatibility decision, not on the adoption decision. Hence they can help understand why there are e.g. different giro-standards, but they cannot be directly used to understand why the US has never fully adopted a giro-system in the first place.[30] Of the compatibility models, those of Economides and Flyer (1998) and Shy (2001) seem most promising, although they will need to be adapted to deal with adoption. In general, the models I reviewed find that the equilibrium outcome(s) depend crucially on the strength of the network effect, with the condition of DePalma and Leruth (1993) providing a useful tool across multiple models.

2.2.3 Role of autarky and spatially separated users

An interesting consequence of network externalities is the occurrence of different (unsponsored) standards in spatially separated regions or countries: why do we speak different languages, why do some countries drive on the left while others drive on the right etc.[31] Several authors have formulated models that analyze how local interaction can lead to local equilibria. For example, Bassanini and Dosi (1998) extend Arthur's (1989) model to an environment with multiple interacting pools of consumers, and find that different outcomes can occur for different pools. Ellison (1993) analyzes adoption by users located on a circle, and finds that strong local interaction can lead to the coexistence of different standards, although a single standard is generally the long run equilibrium (it is an attractor). Cowan and Cowan (1998) analyze adoption of standards by users located on a grid. They show how the interaction of local positive externalities and global negative externalities leads to differences among regions. Cowan and Gunby (1996) apply a similar principle to analyze regional differences in pest-control strategies, while Gunby (1996) applies it to the adoption of the ISO 9000 standard. Puffert (2001) shows how strong local externalities led to local choices for different railway gauges, with smaller networks converting to the gauge of a larger adjacent network as networks became more interlinked.

With the exception of Bassanini and Dosi, all these models assume a homogeneous topology of users (a grid or a circle). This is an important limitation: a homogeneous grid does not contain natural 'niches' that are protected by

[30] Because these models analyze compatibility between standards with equal network effects, it is not possible to apply them to adoption by defining the incumbent technology as just another standard, if the new technology has much stronger network effects.

[31] For an interesting analysis of the language problem, see Shy (2001) and Church and King (1993).

natural barriers/borders and where local standards can thrive. As a result, these models require rather strong conditions for the coexistence of multiple standards. For example the models of both Cowan and Cowan (1998) and Cowan and Gunby (1996) assume global *negative* externalities; with this assumption multiple standards can indeed coexist. In reality, countries provide such natural niches: users interact randomly within a country, but rarely with users abroad. It is obvious that if there is no international interaction, different standards can coexist without negative global externalities. The interesting question is whether this is an equilibrium even if there is more international contact. Since the existing models do not allow for the analysis of this issue, I will build my own extension of the models in chapter 3 and 4 to deal with a setting of semi-autarkic countries.

2.3 Payment Instruments as Networks

Many authors have pointed out the pervasive existence of network externalities in payment systems and their implications for banks and regulators. For example, the June 2003 issue of the Review of Network Economics was dedicated to network effects in payment systems.

2.3.1 Empirical evidence

Empirical studies confirm the existence of network externalities for several payment instruments.

1. *ATMs.* Saloner and Shepard (1995) find that banks with many branches adopted ATMs earlier than banks with few branches. They explain this by the fact that banks with more branches had more potential ATM locations. The larger ATM network of large banks gave the users a larger network benefit (in the early days all ATM networks were proprietary). Paroush and Ruthenberg (1986) analyze ATM adoption by banks in Israel with similar conclusions. Hannan and McDowell (1987) and Sharma (1993) find ATM adoption to be positively correlated with local market concentration, adoption by other banks in the area and bank size; all of these factors point to network effects. Kauffman and Wang (1994) analyze the decision by banks to link their proprietary ATM networks to regional or national 'networks of networks'. Their analysis finds that banks with lots of branches (a proxy for banks with a large proprietary ATM network) joined these networks later. They explain this by the fact

that these larger banks had more to lose (the competitive advantage of their own large network) and less to gain from joining a regional network.

2. *ACH/giro.* Gowrisankaran and Stavins (2002) analyze the decision by banks to adopt ACH technology, using quarterly data on ACH usage by 11,000 US banks covering the period 1995-1997. They find the adoption decision to be significantly and positively correlated with: (1) bank concentration in the banks' micro market (using standard definitions of Metropolitan Statistical Areas or MSA's); and (2) adoption by other banks in that market.[32] Overall they find a moderately large network externality. Adoption appears to be more correlated with adoption of other banks in the same MSA than with the total transaction volume in the market. The authors interpret this as suggesting that the network externality at the consumer level is less important than the externality at the bank level. Ackerberg and Gowrisankaran (2002) use the same data set to econometrically fit a model of the adoption of ACH technology at two levels: the bank and individual customers. It introduces utility and cost at both the bank and consumer level as unobserved variables. The authors find that the bank fixed costs are low; therefore fixed costs cannot explain why ACH isn't used more widely. In contrast, fitted consumer fixed costs are substantial, and a major explanation for the lack of ACH usage. Unfortunately, the estimation method does not allow for a translation of the model estimates to real dollars. Also, the model assumes fixed (unobserved) prices for ACH services across banks; since the model parameters indicate that banks profit substantially if their customers adopt ACH, it remains unexplained why they do not lower prices or subsidize adoption by customers (as was done by European banks).

3. *Debit and credit cards.* Stavins (2001) finds indirect evidence for network externalities in both debit and credit cards: correcting for all demographic differences, usage is highly correlated with usage by others nearby. The same pattern is found by Boeschoten (1998) and Mantel and McHugh (2001).

4. *Cash.* Caskey and St.Laurent (1994) explain the (non-)adoption of the Susan B. Anthony dollar coin by the fact that the coin was unable to gain critical mass in consumer recognition and acceptance. The US is the

[32] Surprisingly there are not only new entrants (banks adopting ACH technology) but also many exiters, suggesting that most of the costs are ongoing, not sunk. As the authors note, part of the exiters may also be banks who did not register any ACH transaction during that quarter.

only developed country without a coin for values over \$0.25. Coins are cheaper than bills for both the central bank (coins have a much longer lifespan) and the operators of vending machines. However, owners of vending machines did not invest in converting machines to accept dollar coins, because they feared consumers would not use them. McAndrews (1997) notes that Canada did succeed in introducing dollar coins because the Bank of Canada withdrew notes from circulation and forced vending machine operators to modify machines.

2.3.2 Theoretical models of payment instruments as networks

ATMs

McAndrews (2003) gives a comprehensive overview and discussion of the theoretical models of ATM network models. I briefly summarize the main models in this paragraph. I already discussed the models of Shy (2001) and Matutes and Padilla (1994) for the compatibility decision for ATM networks. Massoud and Bernhardt (2002) look at the practice of surcharging, where fees (on average \$1.38 per ATM withdrawal) are directly charged to the ATM user by the ATM owner.[33] They show that prohibiting this practice would actually raise ATM prices.[34] Many shared ATM networks are jointly owned by the banks whose ATMs they connect (just as Visa and MasterCard are owned by banks that issue the cards and acquire the merchants). McAndrews and Rob (1996) look at the ownership structure of these networks. They show that shared ownership leads to higher retail prices for ATM services than independent ownership; as a result they argue for greater scrutiny of regulators.[35]

---

[33] In his analysis of bank fees, Hannan (2001) shows how the importance of surcharging has grown: the share of banks that charge such fees has gone from 44.8% in 1996 to 82.9% in 1999, while the average surcharge per ATM withdrawal increased from \$1.19 to \$1.26. These fees are charged by ATM owners and come on top of the fees charges by the bank of the cardholder, on average another \$1.17 for each withdrawal from someone else's ATM. No wonder the PIRG (1999) remarks: "ATMs: Always Taking Money".

[34] The mechanism behind this result is closely related to my own model in chapter 4, partly because the authors use the same spatial differentiation model as I do. Prohibiting surcharging reduces competition and thus allows banks to increase price. Both their and my own result are in line with more general findings of competition under horizontal differentiation as e.g. treated in Anderson, DePalma, et al. (1992).

[35] Interestingly, Rey and Tirole (2000) analyze the cooperative nature of the credit card networks and conclude that it is in the banks' own interest to privatize them. This gives the networks access to capital markets, which is needed to stay competitive.

Credit Cards/Interchange

Credit cards and the related interchange mechanism have been analyzed extensively by both regulators and network economists. This is understandable given the stakes involved. For each credit transaction that a consumer performs at a merchant, the bank of the merchant passes about 1% of the transaction amount to the bank of the cardholder. Since turnover on credit cards in the US alone is approaching $ 1 trillion, the total yearly interchange is in the order of $ 10 billion. Merchants and several regulators charge that this interchange is effectively a form of price-fixing by banks. This argument is forcefully made by, for example, Balto (2000) and Gans and King (2003). Several other models show however, that life is not so simple. Weinberg (2002) looks at the pricing of inter-bank services, such as ATM transactions, Credit Cards or transfers that are 'off-us'. He finds that the nature of competition between banks plays a major role. If banks operate in segmented markets, where each bank has its own natural customer or product base, cooperation of banks in setting inter-bank prices is to be preferred; since each bank has its own monopoly, uncooperative pricing would lead to double marginalization, with higher consumer prices and loss of social welfare as a result. But if banks operate in competitive markets the impact of cooperation in setting inter-bank prices on social welfare is ambiguous and very sensitive to the elasticity of the demand for payment services.

Rochet and Tirole (1999), Schmalensee (2001), and Wright (2001) have all built models to analyze the effect of credit card interchange fees, a payment mechanism between banks that effectively transfers value from the merchant to the cardholder. All these models explicitly take network externalities into account. The crucial issue these models try to answer is whether the interchange mechanism leads to over- or undersupply of credit card payments from a social welfare point of view. A thorough comparison of these models is given by Chakravorti and Shah (2001), who conclude that "the academic literature does not provide a consistent view on the optimal bilateral pricing decisions".[36] The outcomes are highly dependent on the detailed assumptions and specifications of the models.

Money

Perhaps most fundamentally, it has been argued that fiat money itself is a network good, deriving its value from the fact that a critical mass of participants in an economy accepts it as payment.[37] Kiyotaki and Wright (1993) develop

[36] Chakravorti and Shah (2001), p. 6.
[37] Fase (1999) and Shy (2001).

a model where agents trade several commodities among themselves, and show that one of these commodities naturally assumes the role of fiat money. While the Kiyotaki-Wright model focuses on role of money as a transaction medium and store of value, others have used network economics to model the role of money as a unit of account. Dowd and Greenaway (1993) apply the model of Farrell and Saloner (1986) to analyze the competition between currencies. The idea is that consumers build up a frame of reference in terms of prices. This makes switching the unit of account costly, much like learning to use a new measurement system or language. They find that multiple currencies can coexist only if there are positive costs of switching to a different unit of account, because the network benefits will force society towards one single standard. Maintaining dual currencies is socially costly, much like maintaining two languages in a country is costly.[38] An intriguing model is presented by Bak, Nørrelykke, et al. (1999), where agents on a lattice assign value to money using the valuations of their immediate neighbors. Thus the value in equilibrium (the unit of account) is not fixed, and changes in value ripple through the system.

2.3.3 Regulatory implications

The literature has focused on two issues with regard to government intervention and regulation in retail payment systems.[39] The first issue is whether the government should play a role in stimulating their adoption, through subsidies, standard setting, regulation etc. Some like Issing (1999) argue in favor, because network effects may lead to excess inertia in the adoption of socially efficient payment systems. Others argue against such a role, for example because governments tend to pick the wrong technology and standard; and by selecting the wrong standard they may even prevent the adoption of the right standard by the private sector (Gowrisankaran, 1999, on the adoption of ACH systems in the US, and Mantel and McHugh, 2001, on electronic payment networks). Perhaps most outspoken on this topic is Weinberg (1997), who argues that market participants can always reach a sustainable network arrangement,

[38] And indeed, the currency problem as formulated by Dowd and Greenaway is mathematically equivalent to the language problem as formulated by Church and King (1993). Since both apply the same model, they both get the same results: society should only move to one language/currency if switching cost are low enough, and in any case the move should always be to the language/currency with the largest installed base; the value of the Euro should been chosen at par with the DM, Candians should all speak English, and (at a larger scale) the world should adopt Mandarin as its lingua franca.

[39] In addition of course there is the monetary role of government and the prevention of systematic risk in large value payments; both fall outside the scope of this thesis.

provided that side payments or price discrimination is permitted and there are no barriers that prevent market participants from joining other networks. Apart from the heroic assumptions, I find Weinberg's argument and his model somewhat strange; the model does not address the question of how or whether market participants overcome lock-in to reach such a sustainable arrangement.

The second issue is somewhat the reverse: should the government regulate (i.e. restrain) payment networks once they are established? Calls for such regulation can be heard with respect to debit and credit cards. Proponents of such regulation, like Balto (1995, and 2000) and Salop (1990) argue that the increasing returns of payment networks lead to monopolistic power which is being abused by banks. Especially the interchange mechanism is being scrutinized by regulators. The seminal paper is Baxter (1983) who defends the interchange mechanism as being an indispensable enabler of new payment technologies that bring social welfare. As Chakravorti and Shah (2001) conclude in their review of models on interchange, the precise effect of interchange on social welfare is not easy to determine and very sensitive to model specifications. Perhaps as a result, to date regulators have not come to clear point of view: US regulators have wavered (see Chang and Evans, 2000, for a regulatory history of credit cards), the European Commission has recently sanctioned the interchange for cross border debit card transactions (European Commission, 2000), and the Australian regulator is laying the ground for lowering interchange on credit cards and abolishing it on debit cards (Reserve Bank of Australia, et al., 2000).

Summarizing, network effects have been empirically found to exist in ACH transfers and card based payment instruments (ATM, POS debit and credit). Quite a few theoretical models have analyzed the effect of such network effects on the desirability and effect of regulation, generally with inconclusive or even opposing results. Finally, several models have used network effects to analyze the adoption and compatibility decision, notably Shy (2001) and Ackerberg and Gowrisankaran (2002). Shy assumes sponsored standards and I will draw on his model in chapter 4. Ackerberg and Gowrisankaran assume unsponsored standards, but I find their model impractical to analyze the adoption of ACH-transfers in Europe (my objections were analyzed in section 2.3.1 sub 2).

2.4 Conclusions from payment and network literature

The following conclusions can be drawn from the review of existing literature:

1. There is no satisfying explanation for the country differences. Empirical studies find that country idiosyncrasies rather than variables like GDP and crime explain the differences in instrument usage.

2. There is strong empirical evidence that network externalities exist in payment instruments like ACH transfers, debit- and credit card payments and ATMs.

3. The substantial literature on network effects shows that indeed lock-in into an inferior unsponsored standard can occur if the interests of players are not aligned and/or if players have incomplete information about each others' interests. Sponsored standards may reduce, but not eliminate, the occurrence of lock-in. The literature also indicates that spatially separated users may adopt different standards, although the existing models are not readily applicable to a setting of semi-autarkic countries.

3
Adoption and harmonization of unsponsored standards

How come countries that came from a similar background in terms of payment instruments (cash and later checks) end up adopting different payment technologies? In particular why does the US continue to use so many checks even though this is very costly compared to alternatives? And why do European countries have different incompatible giro-systems, making cross-border systems transactions so costly?[1] And finally, why are payment systems national: why do banks within a country use the same instruments, while the choice of instrument varies widely across countries?

To answer these questions, section 3.1 introduces a model to analyze the adoption of an unsponsored network technology by existing firms in an industry (sponsored standards will be discussed in the next chapter). The model is designed to analyze the adoption of ACH/giro transfer payments; this instrument is subject to increasing returns (see Gowrisankaran and Stavins, 2002 and Ackerberg and Gowrisankaran, 2002), and the initial adoption in several European countries took place in an environment where no other network payment systems were present yet. The results can be applied to the adoption of any unsponsored network technology by existing firms.

Sections 3.1 and 3.2 analyze the simplest case, with n firms of different size whose customers transact randomly with each other. The model shows that lock-in is a Nash equilibrium if the industry is fragmented enough. Here lock-in describes the outcome where nobody adopts the new technology, even though adoption by all firms would benefit everybody. Section 3.3 introduces the concept of autarky, where transactions take place disproportionately within firms or within countries. The result of this is that there are less parameter values for which the suboptimal outcome (nobody adopts the new technology) is an equilibrium. Put differently, 'adoption thrives in niches': with autarky even a more fragmented industry will always adopt the new technology.

Section 3.4 analyzes the compatibility decision, if there are multiple incompatible versions of the technology, with a non-zero cost of switching between versions. I find that autarky can have a second effect: the coexistence of multi-

[1] The evidence on US check use and its cost was presented in section 1.4, the evidence on incompatible European giro systems will be discussed in chapter 6.

ple incompatible versions can well be an equilibrium outcome, while the adoption of incompatible standards does not happen if firms or countries transact randomly. Thus autarky is a double-edged sword: it facilitates the adoption of new technologies, but it also can enable the coexistence of multiple incompatible versions.

The model shows that two types of lock-in may well occur: (1) firms in a fragmented industry may fail adopt a new technology, even if full adoption would benefit all firms; and (2) countries may fail to migrate to a common standard, even if migration by all players would raise welfare. Another important outcome of the model is that while the technology landscape can well be heterogeneous *across* countries, it will be homogenous *within* countries.[2] This is an important outcome, because the technology succession model of chapter 8 assumes that technologies are national.

The last section of this chapter summarizes these conclusions in more detail. Table 3.1 in section 3.5 may serve as a useful reference while reading this chapter. It summarizes and compares the equilibrium outcomes of the model under various assumptions.

3.1 Basic model

n banks all face the decision whether to adopt a payment technology g, which is different from their current technology f. They each have a set of customers that make payment transactions using either f or g. The banks have market shares $s_1, s_2, .., s_n$ with $s_1 \geq s_2 ... \geq s_n > 0$.

All banks start from a situation where they are using technology f. Without loss of generality I normalize the cost and benefits per period of technology f to zero, and the total number of customers to one; the number of customers of bank i is then equal to its share s_i. I also assume that the number of transactions per period per customer is fixed (this assumption, as well as the assumption that bank market shares are fixed will be dropped in the next chapter) and normalized to one.

The use of g brings both benefits and costs for the adopting bank and its customers. Let b and c denote these benefits and costs respectively. I define b as benefits per transaction. These benefits could for example be lower costs per transaction as a result of using g instead of f. Alternatively they could represent increased customer fees made possible because of the higher customer

[2] This holds as long as two conditions are met: autarky occurs between but not within countries, and once firms use the same standard, they decide jointly on migration to another standard. In practice both these conditions are met.

value of the new technology g. For simplicity I assume b accrues to the bank of customer that initiates the transaction. c represents the cost per period per customer of using g. In the case of ACH technology for example, this could include the ongoing cost of equipping the customer with transfer forms, educating the customer etc.[3] I assume c is borne by the bank. The distinction between b and c is driven by structure more than the traditional labeling of benefits and costs: c represent per customer investments that have to be recovered through sufficient usage of g. This usage gives net benefits b (difference between extra benefits and costs *per transaction*) for each transaction that uses g instead of the older technology.

I assume the new technology g can only be used for transactions where both the sending party and the receiving party support technology g. If not, the banks and/or their customers have to revert to the old technology f. Many (payment) technologies fit this pattern. For example giro technology can only be used for payments if both parties in a transaction support the technology.[4] I assume that if a bank supports g, all its customers will be able to make or receive payments using g. Let s_g be defined as the joint share of the banks supporting g, and assume that customers initiate their transactions randomly with other customers, i.e. they don't have a preference for transactions with customers of their own bank.[5] The customers of all banks supporting g will then perform a share s_g of their transactions with customers of banks who also support g. The share of g in all transactions is equal to s_g^2, with f being used for the remaining $1 - s_g^2$.

Because the benefits b depend on the usage of g, while the cost c does not, the profit per customer from adopting g is a linear function the usage of g and equal to $s_g b - c$.[6] If g would be used for all transactions in a period (i.e. one per

---

[3] See also footnote 32 in the previous chapter, for evidence from Gowrisankaran and Stavins (2002) that the costs may indeed be ongoing. Their data show that in each period several banks drop the ACH technology, suggesting the cost may be ongoing, not sunk. If the adoption does require per customer investments these should of course be depreciated over time, with c including the depreciation.

[4] For an example outside payments consider Multimedia Messaging Service (MMS), i.e. sending and receiving pictures by mobile phone. This only works if both parties have a phone that supports this activity.

[5] Numbers from the Dutch banking system confirm that this assumption holds for the Dutch market. This assumption does not hold for transaction patterns across national borders, and will be dropped in the next section.

[6] If g is used for a share s_g of a customer's transactions, and if a customer makes one transaction per period, the per customer benefits are $s_g b$ and profits are $s_g b - c$ per customer per period. P. Swann (2002) formulates two conditions for linear network benefits: random transaction patterns and identical transaction usage for each user. Both conditions are met under the assumptions of my model.

customer), the total increase in profit would be $b - c$ per customer. I assume $b > c$, so g is profitable if all banks adopt it.

The structure of b (per 'g-transaction') and c (per period) leads to a network externality. If bank i adopts g it creates benefits of $s_i s_g b$ for the other banks that have already adopted g; here s_i is the share of bank i and s_g is the share of the other banks that already use g. To see this, note that the other users of g can now use g for an extra share s_i of their transactions, so the increase in the share of g-transactions for all other banks combined is equal to the product of these two shares: $s_g s_i$. Since they already incur the cost c for their customers, the externality for all other banks combined is equal to $s_i s_g b$.[7]

3.2 The adoption decision

This section uses the above model to analyze the decision by individual banks to adopt a network technology g. I find two factors to play an important role: (1) the industry structure, in particular the market share of the largest bank or group of banks taking a joint adoption decision; and (2) the availability of alternatives to adopting g, such as upgrades to the existing technology f.

3.2.1 Concept of critical share and role of industry structure

Suppose no bank supports g. If a bank i with share s_i would adopt technology g, it could use g for a share s_i of its transactions. Its incremental profit per customer would then be equal to $s_i b - c$. Bank i will obviously adopt g if $s_i > \frac{c}{b}$. It is also obvious that all other banks will then follow, since they get incremental profits per customer of $(s_i + s_g)b - c > s_i b - c > 0$. If certain banks take their adoption decisions jointly, then these banks act as if they are one player, and we should consider the joint share of these banks.[8] $\frac{c}{b}$ is the *critical share*: if at least one player has a market share larger than this critical share, all players will always adopt g. Let s_c denote this critical share. If transaction patterns are random and f and g are the only options, this critical share is equal to $s_c = \frac{c}{b}$.

---

[7] The other way to derive this result is to note that by adopting g, bank i increases total transaction share of g from s_g^2 to $(s_g + s_i)^2$, an increase equal to $s_i(2s_g + s_i)$; since bank i gets $s_i(s_g + s_i)$, the remaining $s_i s_g$ is an increase in 'g-transactions' for the other banks.

[8] Throughout this chapter and the next one, I will use the word player to describe one or more banks that decide jointly on adoption and compatibility. For an example of such a player, consider the German savings banks; they decide jointly on their payment technology.

Why lock-in cannot occur if $s_1 > s_c$

Adoption by all banks raises every bank's profit per customer by $b - c > 0$. Non-adoption is therefore welfare suboptimal. Throughout the remainder of this chapter I will use the term lock-in to describe such a situation where it would better from a welfare perspective if all banks adopted g, yet nobody does it. As the next proposition states, lock-in can only occur if $s_1 < s_c = \frac{c}{b}$.

Proposition 3.1 *"Lock-in cannot occur in sufficiently concentrated markets": a welfare suboptimal Nash equilibrium exists if and only if* $s_1 < s_c = \frac{c}{b}$, *i.e. the market share of the largest player is lower than the cost/benefit ratio of the new technology* g.

Proof. Nash equilibrium requires that no player can improve his profits through unilateral action. Let s_g denote the joint share of all players that have adopted g. Since by definition $b > c, s_g = 1$ (all banks adopt g) is always a Nash-equilibrium: unilateral deviation (dropping or not adopting g) from this equilibrium will lower any bank's profit from a positive number to zero.

Now let s_1 be the market share of the largest bank. Then there is a second equilibrium where $s_g = 0$ (no bank adopts g) if and only if:

$$s_1 b - c \leq 0 \Leftrightarrow s_1 \leq s_c = \frac{c}{b}. \tag{3.1}$$

This is a Nash-equilibrium because if (3.1) holds for the largest player it automatically holds for all other players. This equilibrium is welfare suboptimal: if all banks adopt g, they all have per customer profit of $b - c > 0$; if no bank adopts g, each bank has per customer benefits of zero. Finally, $s_g = 0$ and $s_g = 1$ represent the only possible Nash-equilibria. To see this, consider the situation where $0 < s_g < 1$, i.e. some but not all banks have adopted g. Then if $s_g < \frac{c}{b}$ the profit per customer to each bank that uses g is $s_g b - c < 0$, so each one of these banks would be better of by dropping g. However if $s_g \geq \frac{c}{b}$ then each bank i that does not use g would have adoption profit per customer of $(s_g + s_i)b - c > s_g b - c \geq 0$, because $s_g \geq \frac{c}{b}$. Thus $0 < s_g < 1$ cannot be a Nash-equilibrium. ■

The crux of proposition 3.1 is that lock-in cannot occur if at least one player gains sufficiently from unilaterally switching to g. The largest player (with share s_1) has most to gain, and he will always switch if $s_1 b - c > 0$, which is the same as $s_1 > s_c = \frac{c}{b}$. If the market leader has sufficient share and switches, all other players will follow suit, because each player can now use the new technology g for a share of its transactions equal to s_1, plus its own share.

Figure 3.1 illustrates where lock-in can never occur. The horizontal axis represents $\frac{c}{b}$, the cost/benefit ratio of g under full adoption. The vertical axis

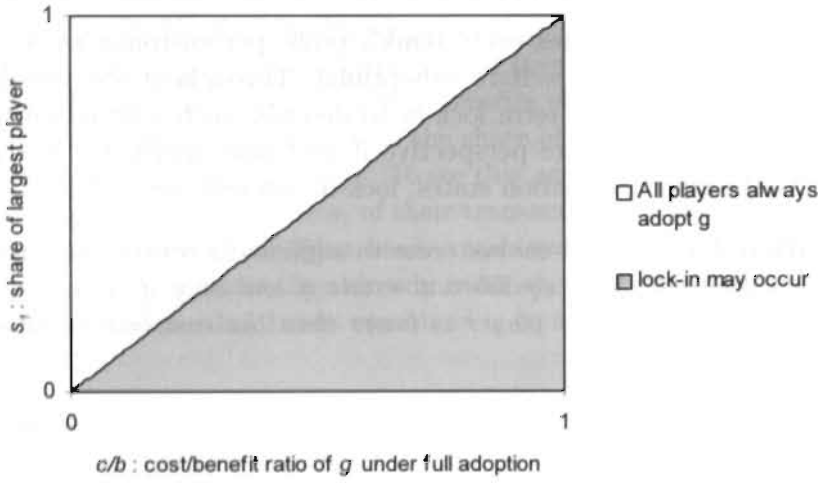

Figure 3.1 Situations where lock-in may occur as a function of the cost/benefit ratio $\frac{c}{b}$ of a new network technology g, and the market share of the largest player

depicts s_1, the market share of the largest bank. The diagonal corresponds to s_c, the critical share which is equal to the cost/benefit ratio of g. Lock-in cannot occur in the upper-left triangle. For example, if g has a cost/benefit ratio of 50%, the necessary critical share, s_c is 50%; lock-in cannot occur if a bank, or group of banks working together, controls more than half the market.

Why lock-in may occur for $s_1 \leq s_c$

Proposition 1 proves that lock-in, or excess inertia, cannot occur if $s_1 > s_c$. What if $s_1 \leq s_c$? From a pure game-theoretic perspective one could argue that even then each player will adopt g. As discussed in the review of the literature on unsponsored standards (section 2.2.1), Farrell and Saloner (1985) prove this. Their functions $B_j(S, X)$ and $B_j(S, Y)$ correspond to my model as follows:

$$\begin{aligned} B_j(S,X) &= 0, \text{ for all } j, S \\ B_j(S,Y) &= s_g b - c, \text{ for all } j \text{ and } S \ (s_g \text{ is the share of all banks in } S). \end{aligned}$$

It is easily verified that their conditions hold in the context of my model:

$$\begin{aligned} B_j(N,Y) &= b - c > 0 \\ B_j(S,X) &\leq B_j(S\prime,X) \text{ for all } j \in S \subseteq S\prime \text{ (both are equal to 0)} \\ B_j(S,Y) &= s_g b - c \leq s_{g\prime} b - c = B_j(S\prime,Y) \text{ because } s_g \leq s_{g\prime} \text{ if } S \subseteq S\prime. \end{aligned}$$

Therefore, their result that lock-in ("excess inertia") cannot occur holds as well. The logic runs as follows.

(i) Suppose all players except i have adopted g. Then player i will also adopt g, because by doing so it raises s_g to 1, and adopting g would raise its per customer profit by $b - c$ which is positive by definition.

(ii) If all banks except i and j use g, bank j will adopt g because it knows that once it does so, bank i will also adopt g as per argument (i), after which bank j will see per customer profit increase by $b - c > 0$.

(iii) Working backwards using induction, it is clear that even if no bank yet uses g, any bank will adopt g since it knows that all players will follow.[9]

There are several arguments why the above reasoning may not apply to most real-life cases.

First, as Farrell and Saloner themselves show, the reasoning only works if players have coinciding interests and complete information about these interests. In practice, neither of these may hold. In the case of giro adoption by the Dutch banks, for example, there was substantial disagreement between two camps over the layout of the customer transfer forms: should they be flexible paper-bases or stiffer punch cards?[10] Information may be incomplete because a bank may not know the other banks' costs and benefits of adopting g. Even worse, a bank may not know its own benefits and costs of adoption with any precision. For the early adopter of g there is always a risk that other players come up with their own version of g and/or that some players choose not to follow the early adopter. Even a remote prospect of this may cause the 'inductive adoption chain' to break down.

In the second place, the track record of these induction proofs is at best mixed if the number of players is large. For example Luce and Raiffa (1957) prove by induction that the equilibrium strategy in a 100-fold prisoners dilemma is non-cooperation for both players on all 100 plays; this outcome is very different from the one observed in actual experiments.[11] While these experiments

[9] Note that this argument does not contradict that non-adoption by all is Nash equilibrium. Nash-equilibrium requires that no player can *unilaterally* improve his outcome. However the inductive argument anticipates moves by *other* players to increase the profit of the player under consideration.

[10] Each party had good reasons to want its own solution. One of the reasons why the Dutch banks created their own giro system in 1966 is that they did not want to confront their customers with transfer forms that were card board punch cards, preferring optical scanning of preprinted forms instead. The existing giro system (Postgiro) on the other hand was not willing to change its system, given the investments made in the older technology.

[11] Luce and Raiffa (1957), p. 98.

show that 2 payers may cooperate even when game-theory predicts they won't, other experiments show that 5 or more players generally fail to cooperate even if game theory says that they will. For example Huck, Normann, et al. (2001) conduct a series of experiments studying oligopolies with two, three, four and five firms in a unified frame. With two firms they find some collusion; three firms tend to produce at the Nash level; markets with four and five firms are never collusive. They conclude that "two are few and four are many".

Third, the externality can cause complications. As was shown earlier, a later player i that adopts g creates an externality $s_g s_i b$ for the others. Player i may try to extract some of this value from the others. This prospect may make it less attractive for the others to adopt g in the first place.

In summary, there is reason to believe that if $s_1 \leq s_c$ lock-in *may* occur, but of course that does not mean it always *does*. The above arguments suggest that lock-in may be an especially relevant phenomenon if: the number of players is large, the new technology has uncertain costs and benefits, and/or the new technology involves standards with lots of (debatable) specification choices.

3.2.2 Effect of upgrading an old technology

Excess inertia can be further enhanced by the existence of an upgrade F to the existing technology f. An example of F could be check verification at the point of sale.[12] The effect of such upgrading is to *raise* the critical share, i.e. the level of participation that is needed for the unilateral and profitable adoption of g.

Proposition 3.2 *The availability of upgrades increases* s_c *for all combinations of* b *and* c.

Proof. See appendix. ■

Figure 3.2 illustrates how upgrading can increase the area where lock-in may occur.[13] The axes are the same as figure 3.1; The critical share is no longer the diagonal, but the broken line above the diagonal. Without upgrades lock-in can occur only in the lower-right triangle. The availability of upgrades enlarges this area adding the flat triangle above the diagonal.

The crux here is that F is compatible with f, so it can be used for all transactions, where g can only be used if both parties have adopted g. So even small players will adopt F unilaterally. At the same time adoption of F

[12] Chakravorti and McHugh (2002) calculate that an unverified check costs a merchant $3.00 per $100 sales volume, compared to $0.60 for a verified check; the authors estimate that 75-97% of checks written at the point of sale are verified at a cost of 2-20 cents per check.

[13] For the purpose of the graph I have assumed that $b_F = \frac{1}{2}b$ and $c_F = \frac{1}{2}c$.

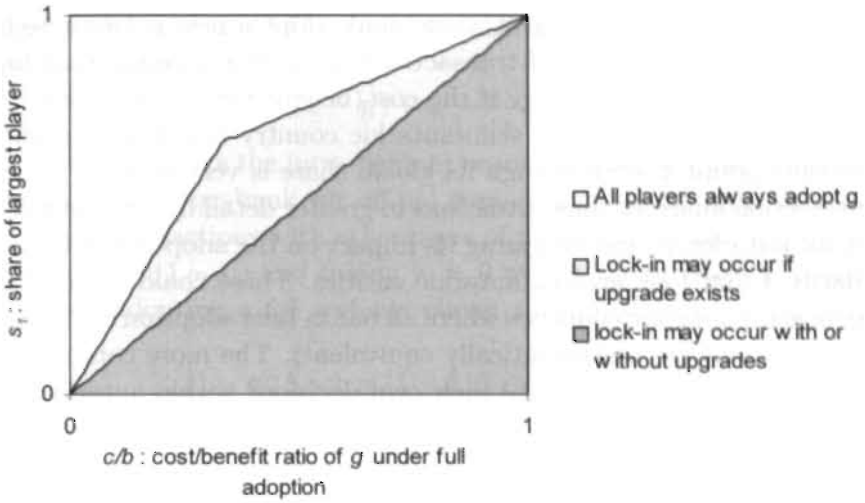

Figure 3.2 Situations where lock-in may occur if upgrades are available
Note: for the purpose of the graph I have assumed that $b_F = \frac{1}{2}b$ and $c_F = \frac{1}{2}c$.

reduces the benefits of g over f, thereby increasing the market share needed to unilaterally adopt g.

Thus a country (or region) can get 'locked-in' to an old technology f. Perhaps the most famous example of such 'extended play' of an old technology is the clipper sailing ship of the late 19th century, where innovations like steel hulls and extra masts (some of the later clippers had seven masts) extended the life of sailing technology by as much as 30 years.[14]

3.3 Autarky and the adoption decision

The model in the previous section assumes an economy where consumers initiate their transactions randomly, i.e. they don't have any preference for transactions with customers of their own bank or within their own country. Reality is more complicated, with a disproportionate number of transactions taking

[14] See C.K. Harley (1973) for the sailing ship effect. In the payment system practice the lock-in effect is further enhanced by the fact that two systems have to be maintained for f and g, each with their own fixed costs; for simplicity our model assumes these fixed costs (other than the fixed cost per customer for g) to be negligible. To the extent that they are non-negligible they will enhance the lock-in effect.

place within institutions and within countries. Intuitively, this should make it easier for institutions to 'go it alone' and adopt a new network technology on their own. If 90 percent of transactions are within a bank, that bank can profitably adopt the technology if the cost/benefit ratio is less than 0.9. Similarly, a large bank in a small semi-autarkic country may find it profitable to unilaterally adopt g, even though its global share is very small.

This section analyzes these situations in greater detail by introducing a parameter for autarky, δ, and analyzing its impact on the adoption of unsponsored standards. I first look at semi-autarkic entities. These could be semi-autarkic banks or semi-autarkic countries where all banks take adoption decisions jointly (these situations are mathematically equivalent). The more complicated situation where multiple banks take their own decisions within autarkic countries is treated in paragraph 3.3.2.

3.3.1 Autarkic banks

If customers initiate their transactions randomly with other customers, the share of intra-bank (or on-us) transactions within each bank i is equal to its market share s_i; the share of transactions with customers of other banks (inter-bank or off-us transactions) is equal to $1 - s_i$.

I now assume instead that for each bank i, the share of a bank's transactions that are inter-bank is proportionally lower. Let q_i denote inter-bank transactions as a percentage of all transactions of bank i, then $q_i = \delta(1 - s_i)$. Here $\delta = 1$ corresponds to complete 'openness': random traffic across banks (the assumption used earlier) and $\delta = 0$ corresponds to complete 'autarky' (no inter-bank traffic).

In practice, within a country such preference for transactions with customers of the same bank appears to be limited; for example within the Netherlands, δ is close to 1.[15] However, the effect is quite significant across countries. Here the basic unit is not banks but countries. s_i is the share of each country in the total, and q_i is the share of international transactions for country i. There is a very strong tendency for people to transact with people in their own country: δ is around 0.02 for traffic across (European) countries.[16] Since for a small entity $q_i = \delta(1 - s_i) \approx \delta$, this means that only 2% of all transactions of a small country are cross border, while for larger players it will be even less.

Proposition 3.3 *"Autarky promotes innovation": a low δ will lower s_c, the critical share needed for the unilateral adoption of* g.

[15] Figures on this are confidential, but people in the industry confirm the pattern.
[16] This figure is derived in chapter 6.

Proof. If the total share of other banks that use g is s_g, then any bank i considering the adoption of g would be able to use it for a fraction of its transactions equal to:

$$(1 - q_i) + q_i \frac{s_g}{1 - s_i}.$$

The first term represents the intra-bank or on-us transactions; the second term corresponds to the inter-bank (or off-us) transactions: g can only be used for the inter-bank transactions with other users of g, and their share is $\frac{s_g}{1-s_i}$. After substituting $q_i = \delta(1 - s_i)$ and taking $s_g = 0$ we get the increase in profit per customer from adopting g for a single player i if nobody else uses g. This is equal to:

$$(1 - q_i)\, b - c = [1 - \delta\,(1 - s_i)]\, b - c.$$

Since this profit is increasing in s_i we can focus on s_1, the share of the largest player. Now lock-in can only occur if profit from adoption by the largest player is zero or negative:

$$\begin{aligned} [1 - \delta\,(1 - s_1)]\, b - c &\leq 0 \Leftrightarrow \\ s_1 &\leq (\frac{c}{b} - 1)\frac{1}{\delta} + 1. \end{aligned} \tag{3.2}$$

By replacing s_1 with s_c in (3.2) and turning the inequality into an equality, we get a new expression for the critical share:

$$s_1 \leq s_c = (\frac{c}{b} - 1)\frac{1}{\delta} + 1.$$

For $\delta = 1$ we get the familiar $s_c = \frac{c}{b}$. Because $\frac{\partial s_c}{\partial \delta} = (1 - \frac{c}{b})\frac{1}{\delta^2} > 0$ (because $0 < \frac{c}{b} < 1$ and $0 < \delta < 1$) the critical share decreases with a lower δ. Hence autarky (a lower δ) *decreases* the critical share and thus the area where non-adoption is a Nash-equilibrium.

Finally, like before, adoption by some but not all players is not a Nash-equilibrium. Suppose the largest player has profitably adopted g, and $s_g = s_1$. Another player j will adopt if:

$$(1 - q_j) + q_j \frac{s_1}{1 - s_j} > \frac{c}{b} \Leftrightarrow 1 - \delta[1 - s_j - s_1] > \frac{c}{b}.$$

This is always the case since $1 - \delta[1 - s_j - s_1] > 1 - \delta[1 - s_1] > \frac{c}{b}$; the second inequality holds because player 1 profitably adopted g. ■

Intuitively, autarky increases the share of 'on-us' transactions, making it more attractive to unilaterally adopt g. In the extreme case of complete autarky ($\delta = 0$), a player will always adopt g since all its transactions are on-us and profits from adopting g are $b - c$ which is positive by assumption. Figure 3.3 shows the critical share for $\delta = 0.5$: compared to figure 3.1 the line representing s_c is tilted downward, and the shaded area, where lock-in may occur, is halved.

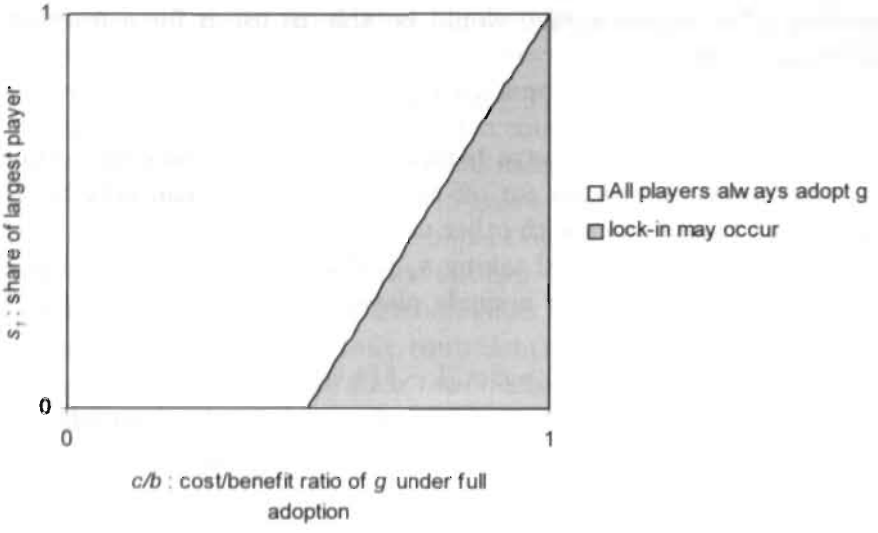

Figure 3.3 Critical share with semi-autarky: $\delta = 0.5$

3.3.2 Multiple players in semi-autarkic countries

In this subsection I explore the situation where there are multiple countries subject to some autarky ($\delta < 1$ *across* countries), and several banks within each country. I assume that $\delta = 1$ *within* countries. The situation now becomes more complex, since for certain market structures there is now a third type of equilibrium: some countries adopt g while others don't.

Let there be m countries, with *global* market shares $s_1 \geq s_2 .. \geq s_m > 0$, and within each country n banks with *national* shares $r_{i1} \geq r_{i2} .. \geq r_{in} > 0$, where:

$$\sum_{j=1}^{n} r_{ij} \quad 1 \text{ for all } i \ (r_{ij} \text{ is the share of each bank within its own country})$$

$$\sum_{i=1}^{m} s_i = 1 \text{ (the share of the countries sum to 1)}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} s_i r_{ij} = 1 \ (s_i r_{ij} \text{ is the global share of bank } i \text{ in country } j).$$

Proposition 3.4 *If there are multiple firms within semi-autarkic countries there are three equilibria:*

(a) *Adoption by all banks. This is an equilibrium for all market structures and all* $0 < \frac{c}{b} < 1$ *and* $0 \leq \delta \leq 1$.

(b) *Adoption by no bank. This is an equilibrium if* $r_{i1} < \frac{c}{b}\frac{1}{[1-\delta(1-s_i)]}$. *The parameter space where this may happen increases as* δ *gets closer to 1, and as* s_i *gets closer to 0.*

(c) *Adoption by all banks in some countries and by no banks in other countries. This can occur for any* $\delta < 1$, *if* $\frac{c}{b}$ *and the market structure meet certain criteria. As* δ *goes to* 0, *this is an equilibrium for any market structure where the largest player in some countries is larger than* $\frac{c}{b}$ *while in other countries it is smaller than* $\frac{c}{b}$.

(d) *Adoption by some but not all banks in a country is not a Nash equilibrium.*

Proof. See appendix. ■

Of these 3 equilibria, the first (all adopt g) is optimal from a social welfare perspective: compared to the second equilibrium (nobody adopts g) it raises welfare per customer by $b - c$. The third equilibrium leads to an increase in social welfare per customer of:[17]

$$W_s = [1 - \delta(1 - s_a)]\, s_a b - c.$$

Here s_a is the share of countries that have adopted g. Note that the social welfare of this third equilibrium is always below the first, so it is indeed a suboptimal equilibrium.

While complicated, applying the model to a situation of multiple banks in semi-autarkic countries yields outcomes that make intuitive sense. The more

[17] This is the increase in welfare compared to non-adoption by all. It is derived as follows. For each country with share s_i where all banks have adopted g, the fraction of transactions where banks can use g is equal to:

$$(1 - q_i) + q_i \frac{s_a - s_i}{1 - s_i}$$

The first term corresponds to the transactions within the country, the second term to the international transactions with other countries that have adopted g. Summing this across all countries that use g gives:

$$\begin{aligned} & \sum_{i \in adopters} \left[(1 - q_i) + q_i \frac{s_a - s_i}{1 - s_i} \right] s_i \\ = & \sum_{i \in adopters} \left[\{1 - \delta(1 - s_i)\} + \delta(s_a - s_i) \right] s_i \\ = & \; [1 - \delta(1 - s_a)] s_a. \end{aligned}$$

autarky there is, the more countries go their own way, with some adopting and others not. Small concentrated countries are the most likely candidates for adopting alone, since larger concentrated countries are more likely to drag the other countries along through cross border traffic.

3.4 The compatibility decision

So far I assumed that only one version of g is available. However, many new technologies come in multiple versions that are incompatible with each other. Sometimes a common standard emerges (as in VHS vs. Betamax and V2000 video formats), but often multiple standards persist. Consider for example languages (Church and King, 1993), and railway gauges (Puffert, 2001).

In the context of our model multiple incompatible versions create a problem for transactions between users of different versions. In practice there are two ways to handle such transactions: (1) migrate to a common version of g; (2) find a specific solution for the inter-standard transactions, such as reverting to another common technology, for example f. Both options bear some cost. This section explores the economics of the trade-off between these options. Not surprisingly I find that autarky plays an important role. Highly autarkic entities will select option (2), while 'open' entities will tend to option (1). One expects different versions of a technology to exist *across* semi-autarkic countries, but much less *within* such countries.

I use the same model as before, except g now comes in incompatible versions $g_1, g_2, ..., g_n$. All these versions have the same adoption economics with per customer benefits of b per transaction, and per customer cost of c per period. Once a version of g has been adopted there are two options to deal with transactions across entities: (1) migrate to a common version of g at a cost of c_m per customer per period, or (2) revert to f, foregoing the benefits b on these transactions. To keep parameters comparable, c_m denotes migration cost per period.[18]

I assume $c_m \leq c$, so migrating to another version of g cannot be more expensive than adopting it from scratch. I also assume that the users of a standard decide jointly on migration to another standard. As the following proposition demonstrates, this implies that different versions can only coexist if there is some autarky.

[18] In practice these migration costs are often one-time costs that have to be depreciated over the following periods.

Proposition 3.5 *If $\delta = 1$ and $c_m \leq c$ multiple versions cannot coexist profitably.*

Proof. Suppose at least two versions of g have been adopted, g_1 and g_2 by entities with joint shares of s_1 and s_2. Without loss of generality, let $s_1 \geq s_2 > 0$. Since g_2 has the smaller share, I only consider the decision by its users to migrate to the larger standard g_1. This is not profitable if:

$$\begin{aligned} s_2 b - c &\geq (s_1 + s_2)b - c - c_m \Leftrightarrow \\ s_1 b &\leq c_m \leq c. \end{aligned}$$

But if $s_1 b \leq c$, technology g_1 is not profitable for group 1, and therefore it is also unprofitable for group 2 (because $s_1 \geq s_2$), violating the assumption that they coexist profitably. ■

The result in this proposition holds if banks using the same version decide jointly on migration.[19] In practice this is often but not always the case. The banks in the largest EU country, for example, generally do not act jointly in payment matters.[20]

As the next proposition shows, things change if $\delta < 1$: different versions can profitable coexist among semi-autarkic countries even if users of the same version act jointly.

Proposition 3.6 *"Autarky enables diversity": for ($\delta < 1$) there are three equilibria:*

(a) *All players use the same version of* g*; this is always a Nash Equilibrium.*

(b) *No player adopts any version of* g*; this is a Nash Equilibrium if the share of the largest player is below* $s_c = (\frac{c}{b} - 1)\frac{1}{\delta} + 1$.

(c) *All players use a different version of* g*; this is a Nash Equilibrium if the share of the largest player is below* $s_c = \frac{c_m}{b}\frac{1}{\delta}$.

---

[19] To see that even for $\delta = 1$, versions can coexist profitably if players with the same version do *not* act jointly, consider the following example: a fragmented market of 100 banks of equal size, with 2 versions of g, each used by 50 banks. If $\frac{c}{b} < 0.5$ both versions are profitable, while an individual bank switch would gain:

$$(0.51 - 0.50)b - c_m.$$

This is negative as long as $c_m > 0.01b$: no individual bank can profitably switch to the other version of g.

[20] I will discuss the German banking sector in more detail in sections 6.2 and 6.3.

Proof. For each semi-autarkic player i the profit per customer is:

$$\begin{aligned} 0 &: \text{if it sticks with } f \\ [(1-q_i)+q_i\frac{s_{gk}}{(1-s_i)}]b-c &: \text{if it } \textit{adopts} \text{ any version } g_k \\ [1-q_i+q_i\frac{s_{gm}}{(1-s_i)}]b-c-c_m &: \text{if it } \textit{switches} \text{ to another version } g_m \end{aligned} \quad (3.3)$$

Here s_{gk} and s_{gm} denote the joint share of players that have adopted the particular versions g_k and g_m, and $q_i = \delta(1-s_i)$ denotes the share of transactions that are off-us.

Part (a). Obviously, the optimal Nash-equilibrium is $s_{gk} = 1$ for any specific version of g_k: all players adopt the same version of g. No player can improve his profitability by unilaterally deviating from this.

Part (b). Following equation (3.2) proposition 3.3, all players sticking with f is a (suboptimal) Nash equilibrium if $s_1 < s_c = (\frac{c}{b}-1)\frac{1}{\delta}+1$, where s_1 is the share of the largest player.

Part (c). First note that each country has to have adopted at least a version of g: if a country has not adopted any version of g it can profitably adopt the version of the largest country that *does* use a version of g (see proposition 3.3). Hence we look for a situation where each entity uses its own version of g (if two entities use the same version we consider them as one). This is an equilibrium if no bank can profitably migrate to an other standard. From expression (3.3) it can be seen that for any entity i, the profit of switching is largest if it switches to the largest standard. Let s_1 be the share of this largest standard, then the following condition must hold for all $i \neq 1$:

$$\begin{aligned} (1-q_i)b-c &\geq [(1-q_i)+q_i\frac{s_1}{(1-s_i)}]b-c-c_m \Leftrightarrow \\ q_i\frac{s_1}{(1-s_i)} &\leq \frac{c_m}{b} \Leftrightarrow \\ s_1 &\leq \frac{c_m}{b}\frac{1}{\delta}. \end{aligned} \quad (3.4)$$

■

Note that if $\delta = 1$, (3.4) becomes $s_1 \leq \frac{c_m}{b}$; and since $\frac{c_m}{b} < \frac{c}{b}$, this means g is not profitable for entity 1, in line with proposition 3.5. Differentiation of right hand term of (3.4) with respect to δ gives the effect of autarky on the critical share:

$$\frac{\partial s_c}{\partial \delta} = -\frac{c_m}{b}\frac{1}{\delta^2}$$

This is negative for all $0 < \delta < 1$: as autarky increases and δ gets lower, the critical share increases and so does the range of parameters (b, c and c_m) for which multiple versions can coexist.

The next proposition considers the welfare effects of such coexistence:

Proposition 3.7 *If $\delta < 1$ there are always some values of* b, c *and* c_m *for which lock-in into different versions of* g *is suboptimal even after allowing for migration costs: all players would be better off if they jointly adopt the largest standard, but no player can unilaterally and profitably migrate to that standard.*

Proof. If all players use the same version of g, overall welfare is equal to $b-c$. If each player uses a different version, the share of g-transactions is equal to:

$$\begin{aligned} & \sum_{i=1}^{n} s_i\{[1-q_i]b-c\} \qquad (3.5) \\ = & \sum_{i=1}^{n} s_i\{[1-\delta(1-s_i)]b-c\} \\ = & [1-\delta(1-H)]b-c, \end{aligned}$$

where $H = \sum_i {s_i}^2$ is the Herfindahl index. If on the other hand all players switch to the version of the largest player, overall profit is equal to:

$$\sum_{i=1}^{n} s_i\,[b-c] - \sum_{i\neq 1} s_i c_m = b-c-(1-s_1)c_m. \qquad (3.6)$$

Thus the net welfare gain of switching to g_1 (the version of the largest player) is equal to the difference between (3.6) and (3.5), which reduces to:

$$\delta(1-H)b-(1-s_1)c_m.$$

This is positive if:

$$\frac{c_m}{b} < \frac{\delta(1-H)}{(1-s_1)}. \qquad (3.7)$$

If s_1 is the share of the largest player, we get ${s_1}^2 \leq H \leq {s_1}^2+(1-s_1)^2$ for all industry structures. By substituting the upper boundary of $H = {s_1}^2+(1-s_1)^2$ into (3.7) and rearranging terms we get a lower boundary for the right hand side of (3.7) equal to $2\delta s_1$. But we know from (3.4) that different versions is an equilibrium if $\frac{c_m}{b} > \delta s_1$. Thus there is always a positive lock-in region $\delta s_1 < \frac{c_m}{b} < 2\delta s_1$ where nobody switches, but overall welfare would enhance if all would adopt the standard of the largest player. ■

The intuition here is that autarky reduces the amount of inter-bank transactions, and thus the profit from adopting a common standard. If there are positive costs to switching to a common version, then these switching costs

may outweigh the extra benefits of being able to use the common version of g also for inter-bank transactions.

Propositions 3.6 and 3.7 have important implications. In many cases banks (or regions or countries) start out as more or less autarkic. For most of the 20th century banks were generally local, and so were transaction patterns; national banks and national companies did not emerge until the latter part of the century. And the internationalization of transaction patterns (and banking) has only barely begun (see the estimate of $\delta \approx 0.02$ mentioned above). This initial autarky fosters the quick adoption of new network technologies. However, due to this quick adoption, the technologies are often in an early stage, without a clear standard. Because of the autarky the welfare loss of incompatible versions is small, so players may prefer speed over compatibility. But if δ subsequently rises (i.e. the banks or countries become more interlinked) the welfare loss may become significant. Even then, it may not be beneficial to migrate to a common standard from an overall welfare point of view, given the migration costs. As δ continues to rise, a point is reached where lack of coordination may become a problem: no individual bank moves, but overall welfare would increase if everybody adopted a common standard, even after allowing for migration costs.

Finally, one may wonder what happens in the most complicated case: multiple semi-autarkic countries, each with individual banks, and technologies $f, g_1, .., g_n$. In that case a fourth type of equilibrium may occur: banks in some countries stick with f, while in other countries banks adopt versions of g that are compatible within the country but incompatible across countries. By combining the results from the previous propositions it is fairly easy to construct such cases; e.g. take a large fragmented country that sticks with f, and two small concentrated countries where the banks each have their own national version of g. Pick a low δ and a relatively high c_m et voilà. Of course this is just mimicking reality in check- versus giro-systems. One large market (the US) sticks to checks and several smaller but concentrated markets (European countries) each have their own version of a giro-system.

3.5 Summary and conclusions

This chapter introduced a model for the adoption of an unsponsored network technology by a set of existing firms. It used this model to analyze possible

TABLE 3.1 Summary of equilibria with unsponsored standards: range of parameter values where they may occur and welfare effects

Equilibrium	Parameter range where equilibrium can occur	Social welfare per customer
1. All adopt same version of g	Always	$b-c$
2. Players adopt incompatible g (only if $\delta < 1$)	$s_1 < \frac{c_m}{b}\frac{1}{\delta}$	$[1-\delta(1-H)]b-c$
3. Some players adopt g, others don't (only if individual firms in autarkic countries)	$l_{lower} < \frac{c}{b} < l_{upper}$	$[1-\delta(1-s_a)]s_a b - c$
4. No bank adopts g	$s_1 < (\frac{c}{b}-1)\frac{1}{\delta}+1$	0

outcomes under a variety of market structures and available technologies. Table 3.1 gives an overview over the results.[21] Overall, I find that:

- Adoption by all firms of the same version of g, while optimal from an overall welfare perspective, is not the only equilibrium outcome. At least three other equilibria may occur, all of these lead to lower social welfare than adoption of the same version of g by all firms.

- In the base case lock-in, where all players stick to the old technology, can occur if the share of the largest player (firm or coalition of firms acting jointly) is smaller than the cost-benefit ratio of the network technology ($\frac{c}{b}$).

- The availability of upgrades increases the critical share. It is now possible that the largest player has a share larger than $\frac{c}{b}$ but all firms adopt the upgrade instead of g.

- Autarky, i.e. a tendency to transact disproportionately within a firm or country lowers the critical share. Autarky thus reduces the parameter space where lock-in into f can occur. However, autarky also enables an-

[21] The formula $l_{lower} < \frac{c}{b} < l_{upper}$ in the table is given by expression (A.6) in the appendix:

$$(1-q_i)r_{i1} + q_i\frac{s_a}{1-s_i} < \frac{c}{b} < (1-q_a)r_{a1} \quad \text{for all } i \neq a.$$

other type of suboptimal equilibrium where firms or countries may adopt incompatible versions of g.

- The most complex (but also most realistic) case of firms operating within semi-autarkic countries can enable a fourth type of outcome: firms in some countries stick with f, while the firms in other countries adopt incompatible versions of g.
- Even in this most complex case, however, firms within the same country always adopt the same technology: f or the same version of g. The model explains therefore why payment systems are national systems.

4

Adoption and harmonization of sponsored standards

The previous chapter showed that the optimal outcome (all firms adopt the new network technology) is generally not the only equilibrium. Industry fragmentation, the availability of upgrades to an older technology and autarky can all facilitate lock-in into an economically inferior technology f and/or into incompatible versions of a new network technology g. A crucial assumption underlying these results is that the technology is unsponsored: there is no owner who controls access to the network. In addition, because firms cannot compete on standards, it was assumed that firms keep the profits of adoption instead of passing them on to the consumer. As a result, demand for transactions was assumed fixed.

Sponsoring can change this. According to economic theory a proprietary or sponsored standard should help to overcome lock-in; proprietary standards allow a firm to 'internalize the externality': if a technology or standard is truly superior, the firm that offers it can expect to capture the benefits of the new technology. The adopting firm can, for example, use the standard to expand market share, charge higher prices, etc. This chapter analyzes whether sponsored standards do indeed reduce the set of circumstances where lock-in can occur.

Somewhat surprisingly, I find that sponsoring does not prevent lock-in if (1) the network effect is small or medium compared to existing firm differentiation and/or customer loyalty; and (2) demand for transactions is relatively inelastic. If these two conditions are met, the results of the previous chapter stand, even with sponsored standards. In fact sponsoring generally *increases* the parameter space for which lock-in can occur. Since there are indications that payment instruments indeed meet these two conditions, this result is quite relevant for the payments industry.[1] In many countries banks indeed cooperate on payment standards, often jointly running clearing houses etc. The regulators generally require that all players have equal access to these systems. One common complaint of larger banks is that common payment systems offer a free ride to

[1] For example Shy derives switching costs for customers of Finnish banks and finds them to be very substantial. Humphrey, Pulley, et al. (1996) find price elasticities for payments in the range of 0.1-0.3.

smaller players. The outcomes of this chapter suggest that regulators should not be overly suspicious of these joint payment systems. It also implies that the fears of large banks are unfounded. They are not providing a free ride to smaller banks; if they were to deny access to smaller firms, their profits would decline due to increased price competition; in fact, regulators may be doing banks a favor by requiring access for all.

A second interesting result is that for a wide range of parameters, sponsored standards too tend to be national standards: players in a country will end up sharing a common version of the network technology, while versions are likely to differ between autarkic countries.[2] Again, sponsoring makes the result obtained for unsponsored standards even stronger; with sponsored standards two semi-autarkic players (firms or countries) may prefer to maintain incompatible standards, even if they could establish compatibility at no cost. The robustness of this result is of interest, since it suggests that (network) externalities indeed often lead to national systems. This may give support to the claim of Dalle (1997), Krugman (1994), Lundvall (1988) and others that technology systems are national. It also suggests that regulators have good reason to press for common standards across Europe: it not only decreases the costs of cross-border transfers, but it may also lead to a more competitive European banking landscape.

Section 4.1 extends the model of the previous chapter to a symmetrical duopoly of two banks and sponsored (proprietary) standards. The more general case of an asymmetric oligopoly (more than two players with unequal market shares) is treated in 4.2. Section 4.3 discusses welfare implications and compares the results of the sponsored model with those of the model for unsponsored standards. Because the game theory, and especially Nash-equilibrium, play a crucial role in deriving the results of this chapter, I analyze their applicability in section 4.4.1. And because my model is similar to that of Shy (2001) I compare the two in section 4.4.2. The last section of this chapter summarizes the results and conclusions.

4.1 The basic duopoly model

My review of the literature on sponsored standards yielded two promising models for analyzing the compatibility decision of firms in the presence of network

[2] This result holds as long as transaction patterns are random across customers of firms in the same country. Clearly two standards can coexist if there are two groups that interact mostly internally and less with each other, for example Apple users (graphics and education) vs. IBM users.

externalities. These were the models of Economides and Flyer (1998) and Shy (2001). Both have advantages and disadvantages. The model of Economides and Flyer allows for the analysis of an oligopoly, while Shy's model deals with a duopoly. On the other hand, Economides and Flyer assume all firms already have a version of the network technology, and the model deals with the compatibility decision, not with the adoption decision. The same holds for Shy, but his model can be more easily adapted to include the adoption decision. Hence I follow the general idea of Shy's model with some important adaptations, such as the inclusion of the adoption decision. In section 4.2 I extend Shy's model to an oligopoly situation (I'll compare our models in section 4.4.2).

The modelling of competition with sponsored standards requires some dampening mechanism, otherwise either all or none of the consumers would flock to one of the standards, and the results become neither interesting nor realistic. Two forms of dampening are used in the literature: consumer arrival and differentiation.[3] The latter approach seems more appropriate in the payment world.

The adoption and compatibility decision of differentiated firms is generally analyzed by assuming firms compete in 2 stages.[4] In stage 1 they independently decide on whether to adopt a new technology and whether to offer compatibility to others. In stage 2 they set prices to maximize profits. The game is then solved backwards: equilibrium profits for stage 2 are derived, and then stage 1 is modeled as a game where the stage 2 outcomes are the payoffs of various adoption and compatibility decisions.

To derive the stage 2 equilibrium outcomes, a model for competition among differentiated products is needed. In their textbook on product differentiation, Anderson, DePalma, et al. (1992) distinguish three types of models for horizontal differentiation between firms: (1) Random Utility Models, in particular the Multinomial Logit (MNL), (2) Representative Consumer Models, in particular the CES, and (3) Address Models, like the one used in Hotelling (1929). They show that the first two models are highly related. This leaves two general approaches: the MNL/CES and address models. Both models have been used

[3] For examples of dampening through consumer arrival (generations or otherwise), see the models of Farrell and Saloner, 1986, and Shy, 1996. For examples of models using differentiation see the models of Matutes and Regibeau, 1988 and 1992 and of Shy, 2001.

[4] Apart from Shy (2001), this approach is also followed by DePalma and Leruth (1993), Jonard and Schenk (1999), Matutes and Padilla (1996). It is also analogous to the location choice models of Hotelling (1929) and Salop (1979) where firms first select a location and then compete on price.

to examine networks sponsored by differentiated firm. I use an address model: the linear city model of Hotelling (1929).[5]

I first describe and analyze the basic model, with network externalities below the critical level defined by DePalma-Leruth, where two incompatible systems can coexist without 'tipping'; in addition I assume fixed transactions demand ($\varepsilon = 0$). I then expand the model to include autarky, i.e. situations where consumers interact disproportionately with other customers of the same firm or country ($\delta < 1$). Finally, I analyze the effect of larger network externalities relative to firm differentiation ($b > t$) and the impact of price sensitive demand ($\varepsilon > 0$).

4.1.1 Base case ($\varepsilon = 0, \delta = 1, b < t$)

Two firms, indexed by $i = 1, 2$, have a fixed location at each end of the unit interval. Consumers are homogeneously distributed along the interval, and purchase one unit of the product from the firm with the lowest price plus transportation costs.[6] As remarked by Hotelling himself, the horizontal spacing of firms need not be physical, but can also represent different points on any attribute dimension valued by consumers. Nevertheless, physical distance may well play a role in banking. Elliehausen and Wolken (1992) demonstrate that banking, and especially maintaining checking accounts, remains incredibly local: in 1989 almost 90% of US households with a checking account held it with a local bank. DeGryse and Ongena (2002) analyze detailed data on 15,000 European bank loans to small businesses; they find that loan rates decrease in the distance between the firm and the bank, while they increase in the distance between the firm and competing banks. The authors propose that the first effect may be driven by superior information of a nearby bank, which allows them to offer lower prices (selecting the better risks), while the second effect is obviously driven by (lack of) competitive pressure. Again this confirms the local nature of (small-business) banking.

In the Hotelling model, prices are net of marginal costs, which are presumed identical for both firms. Equilibrium prices, shares and profits are derived by

[5] The reason to use this model rather than the MNL/CES is that the MNL/CES leads to closed-form specifications of market shares, which makes (numerical) analysis of the Nash-equilibrium very difficult. By contrast, the Hotelling model leads to reduced-form market share functions, greatly facilitating both analytical and numerical analysis. In addition, the Hotelling model can be easily expanded to deal with varying demand elasticity, market structure and coalitions, and network autarky.

[6] Like in the original Hotelling model, Transportation costs may represent the actual transportation costs of covering the distance to the nearest store (firm), or it may represent the disutility of a consumer from buying a product that does need meet his exact needs.

considering the marginal customer with address s_1 on the unit interval $[0, 1]$. This customer is indifferent between buying from either firm, because the price plus transportation cost is the same for both firms. Thus, given p_1 and p_2, the marginal customer (and hence the share of firm 1) is defined by:

$$\begin{aligned} p_1 + s_1 t &= p_1 + (1 - s_1)t \quad \text{(because } s_2 = 1 - s_1\text{)} \Leftrightarrow \\ s_1 &= \frac{p_2 - p_1 + t}{2t} \end{aligned}$$

The second of these equations defines the market share of firm 1 given prices of both firms. Profits for firm 1 are then:

$$\pi_1 = p_1 s_1 = \frac{p_1 p_2 - p_1^2 + p_1 t}{2t}$$

Given the price of the other firm, firm 1 maximizes profits:

$$\begin{aligned} \frac{\partial \pi_1}{\partial p_1} &= 0 \Leftrightarrow \\ p_2 - 2p_1 + t &= 0 \Leftrightarrow \\ p_1 &= \frac{p_2 + t}{2} \end{aligned} \tag{4.1}$$

Since equilibrium is symmetrical, we get $p_1^* = p_2^*$. Substituting this in (4.1) we get the standard Hotelling result:[7]

$$p_i^* = t, \quad s_i^* = \frac{1}{2}, \quad \pi_i^* = \frac{1}{2}t.$$

where:

p_i^*	:	equilibrium prices
s_i^*	:	equilibrium market shares
π_i^*	:	equilibrium profits
t	:	unit transportation cost.

I follow the main assumptions of the model of the previous chapter:

- Firms and their customers use the existing technology f to perform transactions. I normalize the number of customers and the number of transactions to 1.

[7] The model in this chapter requires some juggling of symbols. An overview of all parameters and variables used in this chapter is given at the end of chapter 4.

- A new (network) technology g becomes available to perform these transactions with higher benefits and/or lower costs. However the technology can only be used if both parties support g.

- The transaction patterns are random, so if s_g is the market share of firms that have adopted g, their customers can use g for a share s_g of their transactions.

- There is a per customer cost c of using g while there are benefits of b per transaction (both are assumed to be incremental compared to f). Thus the increase in profits from using g is $s_g b - c$ per customer. I assume $b > c > 0$, so adopting the technology increases profits if $s_g = 1$, i.e. if everybody adopts it.

I also assume that the benefits of using g accrue directly to the consumer, so we can define hedonic prices (corrected for product quality):[8]

$$\hat{p} \equiv p - s_g b. \tag{4.2}$$

Firms play a two-stage game. In stage 1 they independently and simultaneously decide whether to adopt a version of g, and whether to offer compatibility to other firms.[9] *Compatibility can only be established if both parties agree.* Each firm has three options in stage 1: don't adopt; adopt without offering compatibility; and adopt with a compatibility offer. Stage 1 therefore has four possible outcomes: (1) neither firm adopts g, (2) only one firm unilaterally adopts g. (3) both firms maintain incompatible versions of g, and (4) both firms maintain compatible versions of g.[10]

In stage 2 firms set prices and compete for market share à la Hotelling. Perfect foresight and information is assumed. I normalize transportation cost to $t = 1$, so b and c should be thought of as multiples of t. In the case of fixed demand, equilibrium prices can be derived analytically, using the standard Hotelling approach. Table 4.1 shows the results (the formulas are derived in the appendix).

The results in table 4.1 suggest that for any outcome of stage 1, stage 2 will always yield a unique and internal solution. The following proposition confirms that this is true as long as $b < 1$.

---

[8] This is the approach followed in Katz and Shapiro (1985). Alternatively, these benefits can accrue directly to firms (for example in the form of cost savings on transactions), in which cases hedonic prices are not affected by the new technology.

[9] For now I assume compatibility can only be established ex-ante, together with the adoption decision. In many banking and telecom applications it can also be established ex-post. I will examine the consequences of this below.

[10] The fourth outcome occurs if both firms either introduced compatible versions, or offered and accepted compatibility afterwards.

TABLE 4.1 Equilibrium prices, shares and profits of stage 2 for various outcomes of stage 1 of base case competition game between duopolists; $\delta = 1, \varepsilon = 0, b < 1$

Stage 1 outcome	Eq. prices (p_i^*)	Eq. shares (s_i^*)	Eq. profits (π_i^*)
1. Neither firm adopts	1	$\frac{1}{2}$	$\frac{1}{2}$
2. Incompatible versions	$1-b+c$	$\frac{1}{2}$	$\frac{1}{2}-\frac{b}{2}$
3. Only one firm adopts:			
- adopting firm	$1-\frac{b-2c}{3}$	$\frac{1}{2}+\frac{b-2c}{6(2-b)}$	$\frac{(1-\frac{b}{3}-\frac{c}{3})^2}{2-b}$
-other firm	$1-\frac{2b-c}{3}$	$\frac{1}{2}-\frac{b-2c}{6(2-b)}$	$\frac{(1-\frac{2b}{3}+\frac{c}{3})^2}{2-b}$
4. Compatible versions	$1+c$	$\frac{1}{2}$	$\frac{1}{2}$

Proposition 4.1 *"With small or moderate network effects, 2 players will share the market in equilibrium": iff* b<1 *then stage 2 of the game will always yield an internal solution, where both players have positive market share.*

Proof. See appendix for a detailed proof. The essence of the proof is that as long as $b < 1$, the condition for coexistence of incompatible networks, set in DePalma and Leruth (1993), is met. On the other hand, market sharing by two incompatible standards is no longer an equilibrium if $b > 1$; the network benefits then exceed the transportation costs (which I normalized to 1) even for the most distant customer.[11] All customers thus join the largest network and the system 'tips' to the standard of either firm: an increase in the share of a firm increases the network benefits to a user by an amount larger than the increase in transportation cost for the marginal consumer. Instead of an internal solution where two incompatible standards share the market, there are now two corner solutions where either firm captures the whole market. For the remaining analysis in this subsection and the next, I assume $b < 1$. The consequences of $b \geq 1$ are analyzed in section 4.1.3. ■

Stage 1 of the game can now be modelled as a 2 person game, with the 3 options of each firm as strategies and the equilibrium profits of figure 1 as

[11] If $b = 1$ any point between the two firms is an equilibrium if firms adopt incompatible standards, so again there is no unique internal solution.

TABLE 4.2 Payoff matrix for stage 1 of base case competition game between duopolists; $\delta = 1, \varepsilon = 0, b < 1$.

Firm 2 \ Firm 1	No adopt	Incompatible	Compatible
No adopt	$\frac{1}{2}$ / $\frac{1}{2}$	$\frac{(1-\frac{2b}{3}+\frac{c}{3})^2}{2-b}$	$\frac{(1-\frac{b}{3}-\frac{c}{3})^2}{2-b}$
Incompatible	$\frac{(1-\frac{2b}{3}+\frac{c}{3})^2}{2-b}$		$\frac{1}{2}-\frac{b}{2}$
Compatible	$\frac{(1-\frac{b}{3}-\frac{c}{3})^2}{2-b}$	$\frac{1}{2}-\frac{b}{2}$	$\frac{1}{2}$ / $\frac{1}{2}$

payoffs. This is done in table 4.2.[12] As each player has three actions (or 'pure strategies'), there are nine outcomes. However, several of these are identical: if player 1 offers compatibility, but player 2 declines the offer, the outcome is the same as when both players maintain incompatibility; and if only one firm adopts g, it does not make a difference whether that firm offers compatibility or not.

Table 4.2 applies only if both firms decide independently and simultaneously on adoption and compatibility. While this can be realistic in e.g. the software/hardware industry, most banking and telecom applications allow for (in)compatibility to be established after adoption. For example, 2 banks may or may not allow their customers to use the ATMs of the other bank. In that case the compatibility decision becomes a 'subgame' with the four lower-right-hand squares of table 4.2 as payoffs. Since profit under compatibility is always higher than profit under incompatibility ($\frac{1}{2} > \frac{1}{2} - \frac{b}{2}$), compatibility is the unique outcome of this subgame, and table 4.2 can be reduced to the 2 by 2-matrix in table 4.3. This is stated as:

"For small or moderate network effects, competing on standards doesn't pay":

Proposition 4.2 (a) *If 2 players maintain incompatible versions of* g, *the profit for both of them is lower than under any other outcome of stage 1. As a result maintaining incompatible standards is not an equilibrium outcome.*

(b) *If one player unilaterally adopts* g, *profits of both firms go down.*

[12] The upper right hand corner in each square is the profit for firm 1 (whose actions are on the horizontal axis) while the lower left corner is the profit for firm 2.

TABLE 4.3 Payoff matrix with compatibility ex-post

Firm 2 \ Firm 1	No adopt	Incompatible	Compatible
No adopt	$\frac{1}{2}$ $\frac{1}{2}$	$\frac{(1-\frac{2b}{3}+\frac{c}{3})^2}{2-b}$	$\frac{(1-\frac{b}{3}-\frac{c}{3})^2}{2-b}$
Incompatible	$\frac{(1-\frac{2b}{3}+\frac{c}{3})^2}{2-b}$	$\frac{1}{2}-\frac{b}{2}$	$\frac{1}{2}$
Compatible	$\frac{(1-\frac{b}{3}-\frac{c}{3})^2}{2-b}$	$\frac{1}{2}-\frac{b}{2}$	$\frac{1}{2}$

Proof. See appendix. The essence is that for $0 < c < b < 1$ it can be shown that:

$$\frac{1}{2} - \frac{b}{2} < \frac{(1-\frac{b}{3}-\frac{c}{3})^2}{2-b} < \frac{1}{2}$$

and

$$\frac{1}{2} - \frac{b}{2} < \frac{(1-\frac{2b}{3}+\frac{c}{3})^2}{2-b} < \frac{1}{2}.$$

This means that both firms get the lowest profits if they both adopt g and maintain incompatible systems, and the highest if they either adopt compatible versions or none at all. The profits for both firms if just one firm adopts g, fall between these two extremes: $\pi_i^{incompatible} < \pi_i^{one-adopts} < \pi_i^{nobody-adopts} = \pi_i^{compatible}$ for both players (see table 4.2). Therefore the game in table 4.2 and 4.3 has exactly two Nash-equilibria: non-adoption by both firms and full compatibility. 'Incompatible versions' is not a Nash-equilibrium, and neither is 'only one firm adopts' (the adopting firm is better off by dropping g). ■

4.1.2 Semi-autarkic transaction patterns ($\delta < 1$)

Propositions 3.3 and 3.6 in chapter 3 showed the impact if a disproportionate share of all transactions are 'on-us', i.e. they take place between customers of the same firm or country. On the one hand this lowers the critical share needed for unilateral adoption of g, thus reducing the range of parameters where lock-in can occur. On the other hand it enables firms to maintain incompatible versions of g across semi-autarkic firms or countries, as long as there are positive costs of migrating between versions of g. Economically, it would be most efficient to move to the standard of the largest country, since this would minimize migration costs (the largest country doesn't have to migrate). There is an externality in this trade-off: the migration not only benefits the country

making the move, but also the largest country, since it too can now apply the technology for interactions with another country. Hence the question is relevant whether sponsored standards would enable firms to internalize this externality and overcome such lock-in into different versions.

To my surprise, I find the effect running in exactly the opposite direction. For many parameter values, semi-autarky gives firms a commercial incentive to maintain incompatibility, *even in the absence of migration costs.* Thus compared to the unsponsored case, semi-autarkic transaction patterns increase the parameter space where firms maintain different standards.

To analyze the role of autarky in the context of sponsored standards, I again use the Hotelling duopoly model of two firms at opposite ends of the unit interval. However, I now assume the unit interval is divided into two halves ('countries') of equal size. If transaction patterns were random, each customer would perform half of his transactions (payments, phone calls etc.) with other customers in his own country (i.e. his own half of the unit interval), and the other half of his transactions with customers in the other country. Assume now instead, that the two halves are somewhat autarkic: customers on each half of the interval transact mostly with each other. Instead of half their transactions, they perform only a fraction $\frac{\delta}{2}$ of their transactions with customers in the other 'country'. Here $\delta = 1$ corresponds to random interaction, and $\delta = 0$ represents complete autarky, i.e. customers transact only with customers on their own side of the interval. If $\delta < 1$, the semi-autarkic transaction pattern creates a natural border. It turns out that this border softens competition by insulating both players against each other's competition. To see this, suppose firm 1 wants to attract a customer from across this border. Remember that the benefits to a customer depend on what share of his transactions can be made using g. The first 'foreign' customer performs a share $\frac{\delta}{2}$ of his transactions with the other customers of firm 1. However the last 'domestic' customer of firm 1 can perform a share $1 - \frac{\delta}{2}$ of his transactions using g. For $\delta < 1$ we get $\frac{\delta}{2} < 1 - \frac{\delta}{2}$, so the product of firm 1 is worth less to the first cross-border customer than it is to the last customer before the border. As a result, firm 1 can only attract customers form the other side of the border if it makes a downward jump in its price: market shares are no longer a continuous function of prices.[13] The border now acts as an insulator against price competition. For low network benefits, this insulation effect on profits is larger than the gain from compatibility. Proposition 4.3 gives the precise relationship.

[13] And thus the standard Hotelling approach to calculating (Nash) equilibrium prices no longer works. To prove the next proposition I use a concept described by Shy (2001), which he calls Undercut Proof Equilibrium.

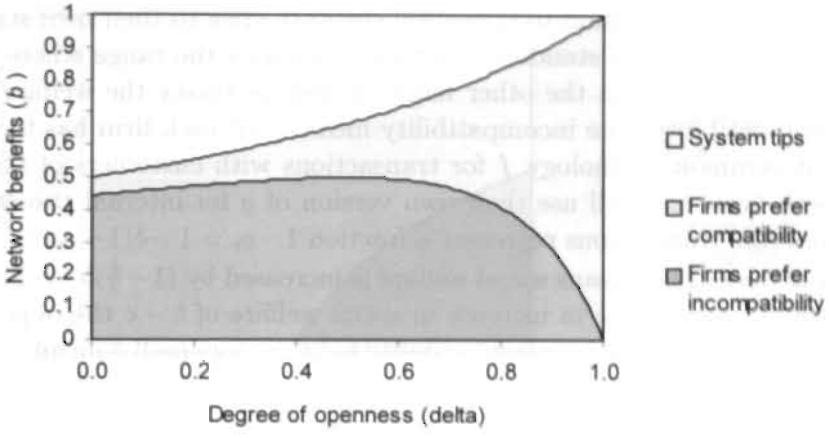

Figure 4.1 Range of values of b and δ where firms prefer (in)compatibility, transaction patterns are not random but semi-autarkic ($\delta < 1$)

Proposition 4.3 *for* $\delta < 1$ *and* b<1, *the outcomes of the game are as follows:*[14]

(a) If $\frac{4(1-\delta)}{(3-2\delta)^2} > \mathrm{b} > \frac{1}{2-\delta}$ *the equilibria are the same as before: either both firms stick with* f *or they maintain compatible versions of* g.

(b) If $\mathrm{b} < \frac{4(1-\delta)}{(3-2\delta)^2}$ *there is only one equilibrium: both firms adopt g but prefer to maintain incompatible versions, even in the absence of migration costs.*

(c) If $\mathrm{b} > \frac{1}{2-\delta}$ *the DePalma-Leruth condition for coexistence of incompatible networks is no longer met, and there are two equilibrium outcomes: (1) both firms maintain compatible versions; and (2) one firm denies compatibility and successfully forces out the other.*

Proof. See appendix. ■

Figure 4.1 gives the critical level of b for each δ. Note that if $\delta = 1$, duopolists always prefer compatibility over incompatibility, in line with the results of section 4.1.1. However, if $\delta < 0.75$, duopolists generally prefer *in*compatibility if network benefits are less than roughly half the unit transportation costs.

[14] c doesn't play a role here. This is because in all equilibria firms either both adopt g (compatible or incompatible) or both stick with f. So costs are the same for both firms. In Hotelling this means that costs disappear from the equations.

Therefore the effect of sponsored standards is again perverse: it gives firms an extra incentive (in addition to migration costs) to stick to their own standard.

As with unsponsored standards, autarky *increases* the range where incompatibility can occur. On the other hand, it also *decreases* the welfare loss of such incompatibility. The incompatibility means that each firm has to use the older but common technology f for transactions with customers of the other firm. Both firms can still use their own version of g for internal transactions. These internal transactions represent a fraction $1-q_i = 1-\delta(1-s_i) = 1-\frac{\delta}{2}$ of all transactions. This means social welfare is increased by $(1-\frac{\delta}{2})b-c$. Since full compatibility would give in increase in social welfare of $b-c$ the opportunity loss in social welfare due to incompatibility is $\frac{\delta}{2}b$. So for small δ incompatibility is no big deal.

4.1.3 Strong network externalities: $b \geq t$

If $b \geq t$, the condition of DePalma and Leruth for an internal stable solution is no longer met.[15] This means that if both firms adopt incompatible versions of the technology, one firm captures the whole market. Without loss of generality let this be firm 1. At first sight this is an attractive prize, especially if $\varepsilon < 1$: in theory firm 1 can now make infinite profits. However the ability to charge usury prices is limited by the threat of entry. I assume firm 2 (or another firm at that location) continues to 'contest' the market, even if firm 1 captures the whole market. The contesting firm could charge a price slightly above 0, to capture at least some share and profit (since the alternative is a profit of 0). Firm 1 can charge a price of no more than $b-1$, otherwise the customers closest to firm 2 will switch (from firm 1 to firm 2) and the market 'tips' the other way. If $p_1 = b-1$, and $\varepsilon = 0$, firm 1 makes a profit of:

$$\pi_1 = (p_1 - c)s_1 = b - 1 - c.$$

Since firm 1 can make profits of $\frac{1}{2}$ by offering compatibility, trying to drive 2 out of the market only makes sense if $b-1-c \geq \frac{1}{2}$. More generally, if $b-c < 1\frac{1}{2}$, the game continues to have two equilibrium outcomes: no adoption and full compatibility. If $b-c \geq 1\frac{1}{2}$ the nature of the game changes. If both firms maintain incompatibility, the market will tip to one of the two proprietary standards. The outcome for an individual firm is no longer certain even if the actions of both firms are known: a firm may either win the whole market or

[15] Formally we have normalized transportation costs to 1 ($t = 1$) so the condition becomes $b > 1$. The original condition is used in the section heading to remind the reader that b should be read as b/t.

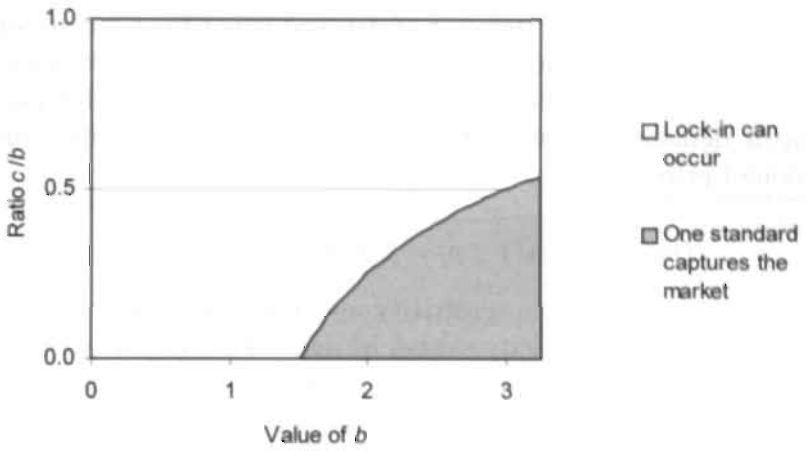

Figure 4.2 Range of c and b where lock-in into old technology can occur if network benefits are large ($b \geq 1$)

be driven out. To derive an equilibrium we would need to know the ex-ante probability that either firms wins a standards war, and some form of risk preference function for the firms.

However, we can say that lock-in into f is *not* an equilibrium; either player can improve his profits through unilateral adoption, after which the other will also adopt and we end up with one of two outcomes: (1) both firms maintain compatible versions, e.g. because they prefer the certainty of profit equal to $\frac{1}{2}$ to the uncertainties of a standards war; and (2) one firm captures the whole market. We can rewrite the condition $b - c \geq 1\frac{1}{2}$ as follows:

$$\begin{aligned} b - c &\geq \frac{3}{2} \Leftrightarrow \\ \frac{c}{b} &\leq 1 - \frac{1.5}{b}. \end{aligned} \tag{4.3}$$

The resulting relationship is shown in figure 4.2.

It follows that for large network externalities, sponsoring does indeed prevent the occurrence of lock-in.

4.1.4 Variable transaction demand ($\varepsilon > 0$)

In the previous sections I have assumed fixed demand (as is usual in both the Hotelling and MNL approaches to differentiation). This leads to the result

that firms can at best maintain their profits by adopting g; all the net benefits flow directly to the consumer. However, if demand is not fixed but sensitive to prices, firm profits under adoption may *go up* compared to the base case of non-adoption, because the lower (hedonic) prices will lead to increased demand. To analyze elastic demand, I assume that consumer demand is a linear function of the (hedonic) price:[16]

$$D(p) = 1 + \alpha(1 - \hat{p}) = 1 + \alpha(1 - p + s_g b). \tag{4.4}$$

Here α is a measure of price sensitivity and $\hat{p}$ is the hedonic price as defined in (4.2). The actual elasticity cannot be derived directly. Remember that in Hotelling p represents the mark-up over (marginal) costs. Assume in equilibrium (where $p = 1$) there is a mark-up of 50% over marginal costs; this corresponds to a contribution margin of 33%. Now elasticity is:

$$\varepsilon = -\frac{\partial D}{\partial p}\frac{(p+2)}{D} = \frac{\alpha(p+2)}{D} = \frac{\alpha(p+2)}{1 + \alpha(1-p)}. \tag{4.5}$$

Since equilibrium price p^* depends in turn on α, we can quantify the relationship: $\alpha = 0.40$ corresponds to roughly $\varepsilon = 1$.[17]

I assume that all customers on the unit interval keep buying from one of the two firms, but the volume that each customer purchases depends on the price according to (4.4). If $\alpha = 0$, we get $D(p) = 1$: each consumer buys one unit of the good (the basic assumption in Hotelling). Note that α does not affect the relative attractiveness of each firm to consumers; the relationship between market shares and prices is the same for all $\alpha \geq 0$.

Compared to the situation with fixed demand, $\alpha > 0$ decreases the parameter range where non-adoption is an equilibrium. Unilateral adoption becomes more attractive if $\alpha > 0$ (compared to when $\alpha = 0$) because the lower hedonic price caused by even partial adoption will increase demand. And once a firm adopts unilaterally, the other firm follows after which both firms can increase profits by establishing compatibility. The shaded area in figure 4.3 shows the parameter space where this process occurs; for these parameters there is only one equilibrium. The following proposition formalizes this result.

---

[16] To keep results comparable across various forms of the model, the demand function is chosen so that for all α we have $D(1) = 1$ and if $\alpha = 0$ we have $D(p) = 1$ for all p.

[17] The value $\varepsilon = 1$ is relevant, because several authors have found price elasticities of payment instruments to be (far) below 1. To derive that $\alpha = 0.40$ corresponds to $\varepsilon = 1$ I use the formulas for p^* derived in the proof of proposition 4.4. Note that $\alpha = 0 \Leftrightarrow \varepsilon = 0$ and $\frac{\partial \varepsilon}{\partial \alpha} > 0$ for all $\alpha > 0$ thus the elasticity is an increasing function of α. Throughout the remainder I will therefore use the term 'positive elasticity' or $\varepsilon > 0$ to mean $\alpha > 0$ and vice versa.

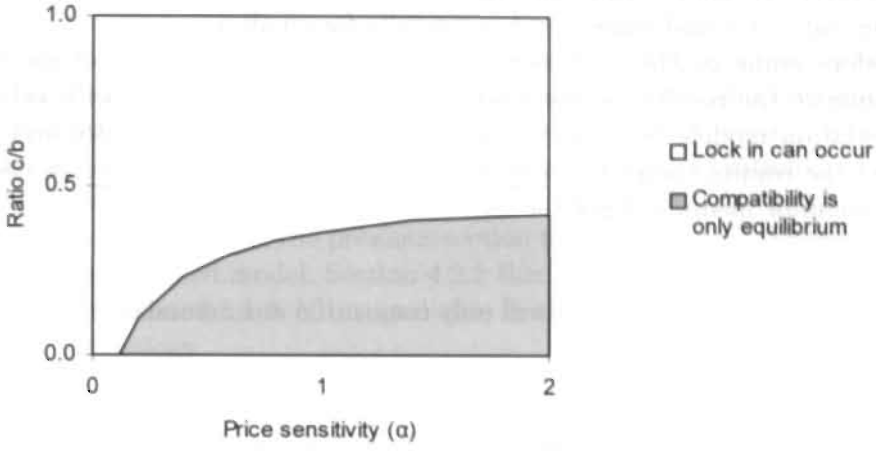

Figure 4.3 Range of price sensitivity (α) and cost-income ratio of g ($\frac{c}{b}$) where lock-in into old technology f is an equilbrium, if transaction demand is variable ($\varepsilon > 0$).

Proposition 4.4 *If* $\alpha \geq 0$, *and* b<1 *(network effects are not too large) there are two equilibria:*

(a) *Adoption of compatible versions. This an equilibrium for all values of* α *and* $\frac{c}{b}$.

(b) *Non-adoption by both firms. This is an equilibrium if* $\frac{c}{b}$ *is high and/or* α *is low, according to the curve in figure 4.3.*

(c) *Neither incompatible versions nor partial adoption (one firm adopts, the other doesn't) form an equilibrium for any value of the parameters* 0<c<b *and* $\varepsilon \geq 0$.

Proof. See appendix.[18] ■

It is interesting to compare these results with those of the unsponsored case. Proposition 3.1 in the previous chapter showed that with unsponsored standards the second equilibrium (neither firm adopts g) can occur if $s_1 < s_e = \frac{c}{b}$. Since in a symmetrical duopoly $s_1 = s_2 = \frac{1}{2}$ this implies that with unsponsored standards the second equilibrium can only occur if $\frac{c}{b} \geq \frac{1}{2}$. However, figure 4.3

[18] Up till now all results have been derived analytically. The results in proposition 4.4 too can be derived for the symmetrical outcomes: neither player or both players adopt. That the asymmetrical state (one firm adopts and the other does not) is not an equilibrium is shown numerically.

shows that the range of parameters where the second equilibrium can occur is larger: all $\frac{c}{b} \geq \frac{1}{2}$ and some $\frac{c}{b} < \frac{1}{2}$, especially for small α.

Before going to the next section (more than two players), let me briefly summarize the results for the model with two players. In line with other differentiation models (e.g. DePalma and Leruth, 1993, Economides and Flyer, 1998) the results completely depend on the strength of the network effect. If it is small or medium, I get the following results:

1. Incompatible standards will only coexist if $\delta < 1$: standards are national.

2. Unilateral adoption is never profitable if $\varepsilon = 0$. The same holds if $\varepsilon > 0$ and $\frac{c}{b}$ is above a certain level (defined in figure 4.3). A first mover then has to anticipate action by his competitor to make it work. If there is uncertainty about this, players may fail to adopt g.

3. Compared to unsponsored standards, the range of parameters where lock-in can occur is not really reduced. In some cases it is even increased.

If the network effect is strong, the results are more in line with the results of e.g. Katz and Shapiro (1986), where the owner of a sponsored standard can internalize the externality and overcome lock-in.

A result that holds for both sponsored and unsponsored standards, is that without autarky, multiple standards cannot coexist: it does not pay to use network standards as a competitive weapon. There are some interesting examples of firms that have tried to compete on standards in payments, but ultimately joined the industry network. For example, Glaser (1988) describes how Citibank unilaterally introduced a proprietary ATM network in 1977 and indeed gained market share. However, they were helped by the fact that it took their competitors 4 years to respond. After that, Citibank held off offers of both national networks (Cirrus and Plus) to join them. In 1991 however, Citibank finally gave in and linked its machines to the national networks.[19] In 1997, the Dutch Postbank introduced its own electronic purse standard, engaging the other banks in a standards war, as both camps tried to sign on consumers and merchants; this failed, and in 2001 Postbank joined the electronic purse standard of the banks.[20]

[19] Kauffman and Wang (1994), p69.
[20] Jongepier (2002).

4.2 Adoption of sponsored standards by an oligopoly

While a (symmetric) duopoly allows for easy analysis, reality is more complicated. There are generally more than two firms in an industry, and they may be of unequal size. These firms may form coalitions that jointly adopt a technology standard at the exclusion of others.[21] Analysis of such situations requires a model of firm competition that can handle multiple firms of unequal size. I have extended the model of the previous section to do just that.[22] Section 4.2.1 describes this extended model. Section 4.2.2 then analyzes competition among multiple differentiated firms of unequal size, in the context of the adoption of a network technology.

4.2.1 A model of competition by an asymmetric oligopoly

I assume that the market consists of a very large number of sub-markets. In each of these sub-markets, there are two competing players. For example, think of a country with 50 cities, where two banks are serving each city. Suppose that initially these were 1-branch banks, but as time wore on these banks clustered into larger chains; I assume this happened randomly, so there may be cities where both branches ended up with same 'chain'. These chains set 'national' prices (same price for all their outlets), and their total production volume is the sum of their market share in each city where they have a branch. I assume the market share in each city is determined by the prices charged by the two branches in that city, using Hotelling's formula. This model could be applied to petrol stations, supermarkets, airlines or any industry where firms compete in many local markets, and where the decision to enter a market can be considered as more or less exogenous (because of restrictions, or because the cost of entering a market is too high, while it is sunk for existing players).

Let me illustrate the concept with a numerical example. Assume there are 50 cities, with in total 100 branches. Suppose bank A ended up with 20 of these branches. Thus A has a 'natural share' of 20%. Let $\hat{s}_A$ denote this natural share. In a random world, one would expect that 20% of A's branches (= 4 branches) are matched against another branch of A; thus A controls both branches in 2 cities. Suppose two other players, B and C, each have half of the remaining branches (40 each), corresponding to a natural share of 40%

[21] Chapter 3 also considered coalitions. It did so by defining s_1 as the share of the largest player, i.e. firm or group of firms deciding jointly on adoption and compatibility of g.

[22] There are existing models that deal with oligopolists in an 'address differentiation' context, for example Salop's circular city (Salop, 1979) and Shy's switching cost model (Shy 2002). However, in both models, firms compete only against their direct neighbors, which greatly limits the analysis of possible coalitions.

($\hat{s}_B = \hat{s}_C = 40\%$). Half of A's remaining 16 branches (i.e. 8 branches) will compete against branches of player B, the other 8 branches of A will compete against player C.

The *actual* share of firm A will depend on it's natural share (number of branches) and the prices charged by the three banks. In the 2 cities where A has both branches it will capture the whole market, in the other cities its share is determined by applying Hotelling's formula to the prices of both competitors in that city.

Following the above notation, let s_i denote the *actual* share of bank i (share of customers), and $\hat{s}_i$ denote the *natural* share (share of branches).[23] The following proposition gives equilibrium prices and shares.

Proposition 4.5 *Given natural shares $\hat{s}_i$, equilibrium is given by:*

$$p_i^* = \frac{1}{(1-k)(2-\hat{s}_i)} \text{ where } k \equiv \sum_{j=1}^{n} \frac{\hat{s}_j}{2-\hat{s}_j} \tag{4.6}$$

$$s_i^* = \hat{s}_i \left(\frac{1}{1-k}\right)\left(\frac{1-\hat{s}_i}{2-\hat{s}_i}\right) \tag{4.7}$$

$$\pi_i^* = p_i^* s_i^* = \hat{s}_i \left(\frac{1}{1-k}\right)^2 \frac{1-\hat{s}_i}{(2-\hat{s}_i)^2}. \tag{4.8}$$

Proof. See appendix. ■

Before I apply this model to network competition, let me give some interesting characteristics of this model.

- *Fits with Cournot.* For the symmetrical oligopoly case, the model leads to results that are directionally the same as the well known equations for Cournot oligopoly: prices and profits are equal to monopoly level ($p = \infty$) if there is only one firm, while they decrease to Hotelling equilibrium ($p = 1$) as the number of firms increases.[24] To see this, note that with n firms and $\hat{s}_i = \frac{1}{n}$ for all i, the equations (4.6) to (4.8) reduce to:[25]

[23] For simplicity I assume marginal costs are equal to 0 for all firms. In fact p_i should be interpreted as the mark-up over marginal cost (as it is in the original Hotelling model).

[24] Because demand is fixed (completely inelastic) monopoly prices are infinite. As $n \to \infty$ prices go to the Hotelling (duopoly) equilibrium because as the number of firms gets very large, we get local duopolies in each city.

[25] This is because for $\hat{s}_i = \frac{1}{n}$ the variable k as defined in (4.6) reduces to: $k = \sum \frac{\frac{1}{n}}{2-\frac{1}{n}} = \frac{n}{2n-1}$, so $\frac{1}{1-k} = \frac{2n-1}{n-1}$ and $\frac{1}{2-\hat{s}_i} = \frac{n}{2n-1}$.

$$\begin{aligned} p_i^* &= \frac{n}{n-1} \\ s_i^* &= \hat{s}_i = \frac{1}{n} \\ \pi_i^* &= \frac{1}{n-1}. \end{aligned}$$

- *Concentration increases prices and profits.* The prices of all firms rise with the concentration of the industry, which is captured by k.[26]

- *Large firms charge higher prices.* The equilibrium price of a firm increases with its natural market share: the biggest firm charges the highest price, providing a 'price umbrella' for the other players. This follows from (4.6). In general, in a market with n firms, any firm with a share larger than $\frac{1}{n}$ sets a price above $\frac{n}{n-1}$, while the smaller firms set a price below that level. What is in effect happening, is that a larger firm charges a relatively high price at the expense of some market share: a larger firm will have a share less than its natural share, while a small firm has a share bigger than its natural share. This is caused by the fact that a large firm has more local monopolies, i.e. cities where it competes against itself.

4.2.2 Possible equilibria

Non-duopoly market structures enable a new type of equilibrium: partial adoption, where some firms adopt g, while others don't. In the unsponsored case as well as in the sponsored duopoly case this could only occur with semi-autarkic transaction patterns ($\delta < 1$). For an asymmetric duopoly adopting sponsored standards it can also occur if $\delta = 1$.Payoff matrix with 3 equal sized firms ($\hat{s}_i = 33.3\%$) and $b = 0.8$ and $c = 0.12$

For example, consider a market with three firms of equal size, and let $b = 0.8$, $c = 0.12$. Applying my model leads to the payoffs in table 4.4.[27] If firms 2 and 3 both adopt compatible versions of g but deny compatibility to firm 1, firm 2

[26] There is an interesting relationship with the Herfindahl index: $k = \sum_{i=1}^{n} \frac{\hat{s}_i}{2-\hat{s}_i}$ and $H = \sum_{i=1}^{n} s_i^2$. In case actual shares are equal to natural shares we get $H \leq k \leq 1$ since $s \leq \frac{1}{2-s}$ (as long as $s \leq 1$ which is by definition the case) . More generally, k is an "allowable" concentration index in the sense of Encaoua-Jacquemin (as defined in Tirole, 1989, p.222), because it: is (1) invariant to permutations in market shares between firms, (2) satisfies the Lorenz condition (preserving the mean, while making the shares more skewed, increases k), and (3) for equal sized firms, k is decreasing in the number of firms. However, the relationship between k and H is not monotonic, i.e. k measures asymmetry in a different way.

[27] These were calculated using the formulas that are derived in the proof of proposition 4.6.

TABLE 4.4 Payoff matrix with 3 equal sized firms and $b = 0.8$ and $c = 0.12$

Firm 2 and 3 \ Firm 1	No adopt	Own version	Compatible with firm 2 & 3
No adopt	0.50 0.50	0.50 0.43	
Own version	0.36 0.51	0.34 0.43	
Compatible with firm 1			0.50 0.50

and 3 each make a profit of 0.51, higher than all other outcomes. Furthermore, there is not much firm 1 can do about this: if it too introduces g, but the others deny compatibility, profits of firm 1 go down from 0.36 to 0.34.[28]

In the unsponsored case, once a player adopts a technology, it cannot deny access to others and all firms end up adopting the technology. In the sponsored case, we can end up with 'semi lock-in': part of the industry adopts, the rest doesn't. This means that we can have three generic outcomes:

1. The largest player (firm or coalition of firms acting jointly) cannot profitably adopt the technology. In that case lock-in into the old technology may occur.

2. The largest player can profitably adopt the technology and no other player can adopt an incompatible version of the technology and improve profits.[29] This leads to the semi lock-in of the earlier example in table 4.4.

3. The largest player can profitably adopt the technology and at least one other player can then profitably adopt an incompatible version. In that

[28] It should be noted that this critically depends on the precise rules of stage 1. If adoption is irreversible, then firm 1 can 'force' compatibility upon 2 and 3: firm 1 has no way back, and now the coalition's best option is to offer compatibility. We are then in the (Stackelberg) realm of subgames, threats, credible commitments etc. For example, if 2 and 3 could commit to maintaining incompatibility even if 1 enters, they could deter entry by 1.

[29] The largest player *will* deny compatibility to the others. Because profit under full adoption is equal to profit if nobody adopts, the first player will only make a unilateral move in the first place if its profit after doing so is higher than its profit under full adoption of compatible versions.

case both players can improve profits by offering and accepting compatibility: universal compatibility is the equilibrium outcome.[30]

4.2.3 *Analysis of four polar market structures*

The combination of industry structures and parameters leads to myriad situations to be analyzed. I therefore analyze four polar cases:

1. Symmetric duopoly. Both firms have a natural share of $\frac{1}{2}$.
2. General duopoly. Two firms have natural shares of $\hat{s}_1$ and $\hat{s}_2 = 1 - \hat{s}_1 > 0$.
3. Gorilla versus a competitive fringe.[31] A large firm has share $\hat{s}_1$, the other $n-1$ firms are equally small, each with a share equal to $\frac{1-\hat{s}_1}{n-1}$. I analyze the situation where the Gorilla acts alone, while the fringe decides jointly on adoption and compatibility ($\hat{s}_2 = 1 - \hat{s}_1$).[32] This case is different from the general duopoly (where we also had $\hat{s}_2 = 1 - \hat{s}_1$), because the second player (the united fringe) only acts jointly on adoption and compatibility. Each of the fringe firms still sets its own price. For a real world example, think of Microsoft versus the unlikely alliance of Sun, Oracle, Netscape etc.
4. Symmetric oligopoly. n firms all have equal share $\frac{1}{n}$. I analyze the case where a coalition of m firms decides jointly on adoption and compatibility, while the other $n-m$ react by forming a counter-coalition that also acts jointly .

For each of these I checked which outcomes are a Nash equilibrium by performing a numerical gridsearch along the following variables:[33]

- $\frac{c}{b}$ going from 0 to 1 in increments of 0.01
- b going from 0.1 to 1 in increments of 0.01; since variable b did not affect the occurrence of equilibria the two charts below were drawn for $b = 0.8$.

---

[30] This assumes that under compatibility profits for both partners always exceed profits under incompatibility. Numerical analysis confirms that this is always the case for the polar cases analyzed in the next section. here incompatibility is meant to describe a situation where all players adopt g, but versions are not compatible; it does not apply to a situation where one player has not (yet) adopted any version of g: in that case profits for some firms may well exceed those under full compatibility (as the example in table 4.4 showed).

[31] The terminology is taken from Perloff (1991) ch. 10: "Where does the Gorilla sleep? Anywhere the Gorilla wants to sleep."

[32] The other cases, where the fringe is unable to act jointly, follow quite easily from this extreme.

[33] The grid search was performed using a Pascal computer program avaiable upon request.

- $\hat{s}_1$ (the share of the coalition taking the initiative) going from 0 to 1 in increments of 0.01.[34] Note that in all cases $\hat{s}_1$ defines the market structure.

- n, the number of firms, which I set at 3, 10, 100 and 1000.

For each set of parameter values I derived equilibrium profits for both players under four outcomes of stage 1: (1) neither player adopts; (2) player 1 adopts unilaterally; (3) both players adopt incompatible versions; and (4) both players adopt compatible versions. The equations that give these profits under these four scenarios are derived in the proof of proposition 4.6.[35]

First, I verified that profits for both players under outcome (4) exceed those under outcome (3): both players improve profits if they establish compatibility. The numerical analysis that this is true for all cases analyzed. Second, I determined the applicability of equilibria by the calculating the following variables:

$$\pi_1^{unilat} - \pi_1^{non-adopt} \tag{4.9}$$

$$\pi_2^{incompatib} - \pi_2^{other-unilat}. \tag{4.10}$$

Here $\pi_2^{other-unilat}$ denotes the profit to player 2 if the other player (player 1) adopts unilaterally. If (4.9)> 0, player 1 has an incentive to deviate from non-adoption by unilaterally adopting g. If in addition (4.10)> 0, the second player can improve profits by also adopting g even if player 1 denies it compatibility. These two conditions now determine the equilibria as follows:

1. Lock-in (where neither player adopts g) is an equilibrium iff: (4.9) ≤ 0 for both players: neither player gains by unilaterally deviating from non-adoption.

2. Semi-adoption is an equilibrium if (4.9) > 0 and (4.10) < 0. Player 1 gains from unilateral adoption, player for at least one player while the second player does not gain from also adopting g if player 1 denies compatibility (which it will, as argued in footnote 29 of this chapter).

3. All firms adopting compatible versions is always an equilibrium unless (4.9) > 0 and (4.10) < 0, i.e. unless semi lock-in is an equilibrium.

---

[34] In the case of a fragmented market I obviously increased $\hat{s}_1$ in steps of $\frac{1}{n}$.

[35] I restrict myself to analyzing unilateral adoption by player 1. Polar cases 1 and 3 are symmetrical, and since in the gridsearch s_1 goes through the full range from 0 to 1 all cases are analyzed. Case 2 (Gorilla) is not symmetrical, and here I indeed restrict the analysis to unilateral moves by the Gorilla, assuming unilateral moves by the fringe are unlikely.

Using this approach I get the following results.

1. Symmetric duopoly. This case has been analyzed in the previous chapter, where is was found that neither firm can profitably adopt the technology while locking out the other.

2. Asymmetric duopoly. Numerical analysis shows that there are no asymmetric duopolies where one firm wins by maintaining incompatibility. It is not profitable for the largest firm to lock out the other. The intuition behind this result is that in an asymmetric duopoly, the large firm charges a high price in equilibrium (without adopting g). Competing on network standards means lower prices, which tends to depress the high profits of the larger firm.

3. A 'Gorilla' against a 'competitive fringe'. Figure 4.4 shows the combinations of Gorilla size and ratio $\frac{c}{b}$ where the large firm can profitably lock-out the rest (999 small players). The axes are the same as in figure 3.1 and again the lower right side shows the area where two equilibria can occur: (1) all firms adopt g and establish compatibility; and (2) lock-in into f, where no firm adopts g. But the area where this happens is much larger than in figure 3.1. And the white area where full compatibility is the only equilibrium is much smaller. In addition there is now an area where semi-adoption is the only equilibrium: the large player adopts g and denies compatibility to the others who then stick to f, rather then adopt their own version of g. However this requires rather extreme parameter values: the share of the Gorilla must be more than 40% and the cost benefit ratio of g must be below 0.35.

The number of firms only plays a minor role if it gets very small. The graph was drawn for 1000 players. For 10 and 100 fringe firms the curve hardly shifts. The results only change if the number of fringe firms gets very small (3 or 4). The curve then shifts to the left, increasing the dark shaded area where lock-in can occur.

4. A fragmented market with two opposing coalitions of small firms. Figure 4.5 shows the combinations of coalition share and $\frac{c}{b}$ ratio that allow a coalition to lock-out the rest, for markets with 1000 equal sized players. The axes and shading are the same as in figure 4.4. The dark shaded area is now back to the lower right triangle, the same as it was in figure 3.1 of the previous chapter. However semi lock-in can now occur for more modest values of $\frac{c}{b}$ and s_1. If slightly more than half the market sticks together it can lock-out the rest, assuming the cost/benefit ratio is in the right range.

The graph was made for a market of 1000 equal sized firms, but the curves hardly shift for 10 and 100 firms. For 3 firms the curve shifts to the upper left, increasing the dark shaded area where lock-in can occur.

In summary I find that for all of the structures analyzed there is never coalition with share $\hat{s}_1 \leq \frac{c}{b}$ that can profitably adopt the technology, and thus

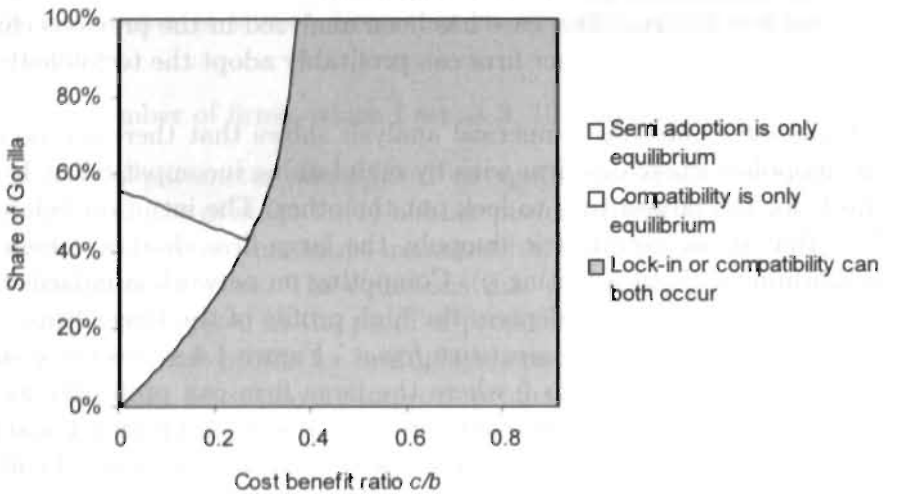

Figure 4.4 Equilibrium outcomes in a market dominated by a large player ('Gorilla')
Note: the graph was numerically generated for the case of a large player agianst a fringe of 999 equal sized small players, with $b = 0.8$.

for all $s_1 \leq \frac{c}{b}$ lock-in can occur. In addition, for each of these structures there are at least some values $\hat{s}_1 > \frac{c}{b}$ where lock-in can occur. This leads to the following proposition.

Proposition 4.6 *Compared to the model for unsponsored standards, the availability of proprietary standards increases the range of parameter values where suboptimal equilibria can occur. This holds for the following industry structures: (i) any duopoly, (ii) any Gorilla vs. competitive fringe, (iii) any number of equal sized firms.*

Proof. See the appendix for the equations that give the profits under general industry structures and coalitions. Figures 4.4 and 4.5 were generated numerically using these equations. They show that for polar cases 3 (Gorilla) and 4 (fragmented market with coalitions) lock-in can occur for all $\hat{s}_1 \leq \frac{c}{b}$ and some $\hat{s}_1 > \frac{c}{b}$. In addition for any duopoly lock-in can occur for all $\frac{c}{b}$. By contrast, with unsponsored standards lock-in can only occur if $s_1 \leq \frac{c}{b}$. ■

The analysis in this section assumes fixed demand ($\varepsilon = 0$). Combining the oligopoly model with price sensitive demand is beyond the scope of this thesis. However, we can make an educated guess on the effects of $\varepsilon > 0$ by extrapolating the results of the previous section (symmetric duopoly). In general, if $\varepsilon > 0$, both unilateral adoption and full compatibility become more attractive

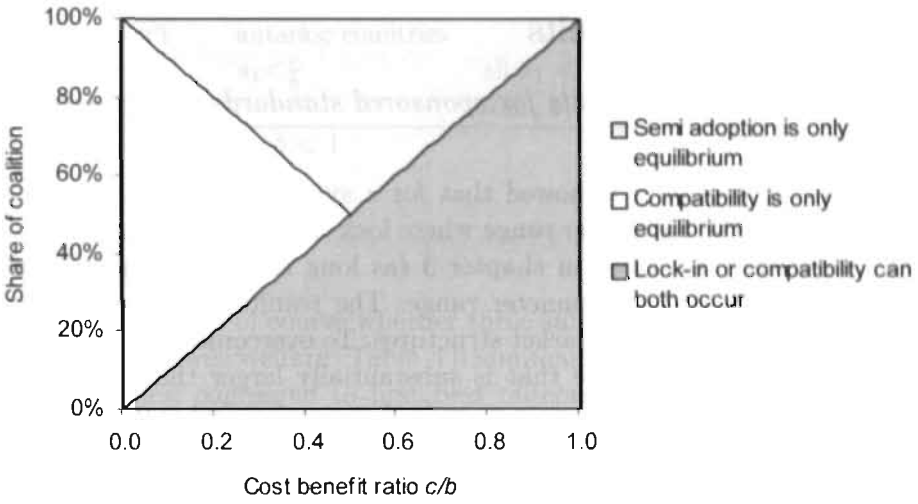

Figure 4.5 Equilibrium outcomes for a coalition of equal players in a fragmented market

Note: the graph was numerically generated for the case of 1000 equal sized players in two coalitions of varying sizes, with $b = 0.8$.

relative to non-adoption. Hence I would expect the curves in figure 4.4 and 4.5 to move *to the lower-right*. The range where lock–in can occur would decrease, and the range where compatibility is the only outcome would increase. But the range where semi lock-in occurs (some players adopt, the rest doesn't) would also increase.

4.3 Discussion of results

4.3.1 Comparison of results for sponsored standards with unsponsored standards

To summarize, section 4.1 showed that for a symmetric duopoly, sponsoring does not reduce the parameter range where lock-in can occur, compared to the unsponsored case described in chapter 3 (as long as $b < t$); on the contrary it tends to increase this parameter range. The results of section 4.2 extend that result to more general market structures. To overcome lock-in, the largest player(s) often needs a share that is substantially larger than that given by the rule for the unsponsored case: $s_1 > \frac{c}{b}$. A more detailed comparison of the results is given in table 4.5. I find that the same four equilibria that can occur with unsponsored standards, can also occur with sponsored standards: (1) all firms adopt the same version of g; (2) firms adopt incompatible versions of g and have to use the old technology f for transactions between customers of firms that use different versions of g; (3) some firms adopt g, while others don't; and (4) no firm adopts g. Outcome (1) is the socially optimal outcome under any specification of the model, and the others are socially suboptimal.

The parameter range for which these suboptimal equilibria can occur is wider with sponsored standards. For example, with unsponsored standards firms will only maintain incompatible versions if there are positive costs of migrating to another version ($c_m > 0$). With sponsored standards two firms can increase profits by maintaining incompatibility even in the absence of migration costs, as long as $b < \frac{4(1-\delta)}{(3-2\delta)^2}$.

The only exception this pattern is the case where $b > 1$ and network effects are strong compared to existing firm differentiation. In that case there is generally only one outcome: one standard captures the market (either offered by one firm that serves the whole market or by two firms that have established compatibility).

TABLE 4.5 Comparison of outcomes with sponsored and unsponsored standards

Equilibrium		Occurrence with unsponsored stds.	Occurrence with sponsored stds.	Effect on area where lock-in can occur
1.	All adopt compatible g	Always	Not if low $\frac{c}{b}$ and large coalition or Gorilla	Sponsoring decreases area
2.	Players adopt incompatible g (only if $\delta < 1$)	Only if $c_m > 0$	Even if $c_m = 0$	Sponsoring increases area
3.	Some players adopt g others don't	Only if $\delta < 1$ and if banks within autarkic countries	Even if $\delta = 1$	Sponsoring increases area
4.	No firm adopts g	$s_1 < \frac{c}{b}$	all $s_1 < \frac{c}{b}$ and some $s_1 \geq \frac{c}{b}$	Sponsoring increases area

Note: table gives results for $b < 1$

4.3.2 Welfare effects

A relevant question is of course whether these suboptimal equilibria really lead to substantially lower welfare. Table 4.6 summarizes the welfare effects of the various outcomes, compared to first best outcome where all firms adopt the same version of g. Depending on the parameter values, these losses can be substantial. Non-adoption is the most expensive, with an opportunity loss of $b - c$ compared to full adoption. The other two equilibria are less costly, each for a different reason. Adoption by just one player can occur only if $s_1 > 50\%$ (more or less), so the loss versus first best is at most $\frac{3}{4}(b-c)$, or three-quarters of the loss in case of non-adoption.[36] Incompatible versions can only occur if $\delta < 1$. That same δ, however, limits the welfare loss; the most costly cases occur where δ is relatively large (say 0.5 to 1). Since for payments δ is very low, in the order of 2% (as was already mentioned in the previous chapter), the actual welfare loss due to incompatibility across countries may be limited.

Table 4.6 holds for cases where transactions demand is fixed, and although the structure of the model is different, the welfare effects are the same as those

[36] This follows by taking the formula for the loss vs. first-best for equilibrium 3 in table 4.6. The loss is equal to:

$$(1 - s_1^2)(b - c) < \frac{3}{4}(b - c) \text{ for } s_1 > 50\%.$$

TABLE 4.6 Welfare effects of different equilibrium outcomes for sponsored standards (fixed demand)

Equilibrium		Social welfare	Loss vs. first best
1.	Compatible versions	$b-c$	0
2.	Incompatible versions (symmetric duopoly)	$(1-\frac{\delta}{2})b-c$	$\frac{\delta}{2}b$
3.	One player with share s_1 adopts, the others don't ($\delta=1$)	$s_1^2(b-c)$	$(1-s_1^2)(b-c)$
4.	No player adopts	0	$b-c$

in table 3.1 in the previous chapter.[37] There is, however, a difference in who profits from adopting g. In the previous chapter I assumed that firms capture the full benefit of g: if all firms adopt a compatible version of g they raise transaction price by b, thus keeping the hedonic price constant. Since they also bear the cost of c, their increase in profit per consumer is $b-c$. In the models of this chapter, by contrast, the full benefit accrues to the consumer: firm profit is the same under full compatibility and non-adoption, while social welfare goes up by $b-c$.

The welfare effects change if transactions demand is price sensitive ($\varepsilon > 0$). It is no longer obvious that full adoption is socially optimal. The price war in case of incompatibility leads to lower prices and higher demand, which greatly enhances consumer welfare. Conceivably this may be more than enough to offset any lower firm profits due to the price competition. If that were the case, incompatibility would be the first best outcome and since incompatibility is not Nash equilibrium (at least for all $b < 1$) this would be a problem for a social planner. As the following proposition demonstrates however, compatibility is the first best outcome under all $\alpha > 0$ and $0 < c < b < 1$.

Proposition 4.7 *For moderate network effects, and all $\alpha \geq 0$, the adoption of compatible versions of* g *by both firms is the socially optimal outcome.*

Proof. See appendix. ■

If demand is price sensitive, the difference in pricing under the sponsored and unsupported regimes is no longer just a socially neutral wealth transfer between consumers and firms. Sponsored standards now lead to lower prices and thus higher volumes per consumer. This makes sponsored standards socially more attractive than unsponsored standards, if $\varepsilon > 0$. A social planner faces a trade-

[37] Note that in a symmetric duopoly $H = \frac{1}{2}$. Also note that s_1 in equilibrium 3 refers to the share after unilateral adoption of g (the outcome is not symmetrical so in general $s_1 \neq \frac{1}{2}$.

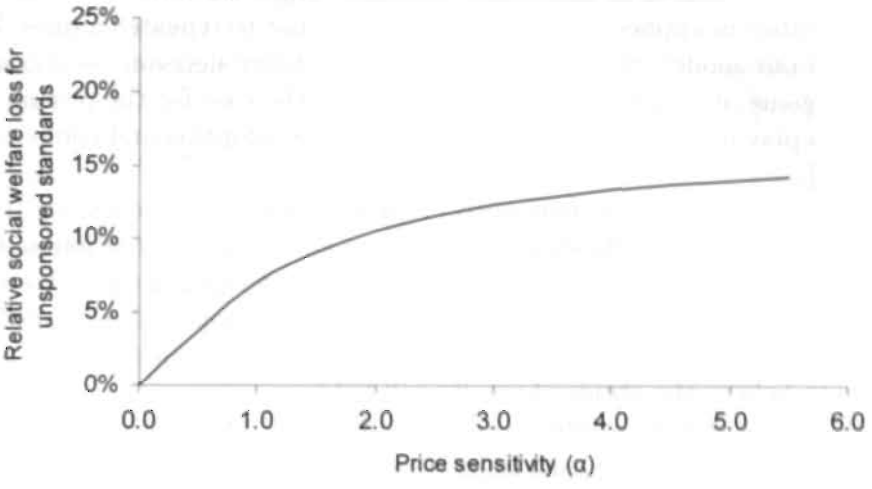

Figure 4.6 Relative loss in social welfare for adoption of unsponsored standard compard to first best outcome with sponsored standards

Note: formally the loss is defined as: $\frac{W_S^{sponsored} - W_S^{unsponsored}}{W_S^{sponsored} - W_S^{non-adoption}}$. The graph was generated using the equations derived in the proof of proposition 4.7, taking $b = 0.3$ and $c = 0.12$.

off: unsponsored standards reduce the occurrence of lock-in, but at the cost of some social welfare.

Figure 4.6 shows the size of this social welfare loss as a function of price sensitivity. The loss here is defined as the welfare gain that is missed by adopting unsponsored standards as a percentage of the maximum welfare gain if sponsored standards were adopted.

To put the results in perspective, the price elasticity for payment services is found to be below 1 by several authors, corresponding to $\alpha < 0.4$.[38] In that case the relative welfare loss is in the order or 5-10%.

4.4 Discussion of model and main assumptions

4.4.1 Does Nash-equilibrium apply?

The results depend crucially on the game-theoretic frameworks used, and in particular on the concept of Nash-equilibrium. Nash equilibrium is explicitly used to solve the stage 1 game, and it is implicit in the assumption that firms

[38] See expression (4.5) for the derivation.

compete à la Hotelling in the second (pricing) stage. However, the concept of Nash-equilibrium applies to one-shot games, and not to repeated games. While stage 1 of our model (the adoption and compatibility decision) is arguably a one-shot game, it is not obvious that this is also the case for the pricing game that firms play in stage 2 (after they have made the adoption and compatibility decisions).

If pricing were indeed a (frequently) repeated game, then according to various versions of the Folk theorem, any outcome above the 'reservation' outcome (or minimax solution) of the single game can form an equilibrium.[39] For example, firms may reach the monopoly outcome (which means infinite prices and profits in the case of perfectly inelastic demand). Nevertheless, I have several reasons to believe the obtained results will hold.

1. Pricing is at best an infrequently repeated game. Once firms have committed to a low price, it is not easy to increase it substantially. Decreasing price is a bit easier; but the alternations of high and low prices that are often observed in experimental oligopoly settings, before players reach 'collusive' outcomes, seem difficult to realize in a practical setting.

2. The outcomes depend on the relative, not absolute, attractiveness of the various pricing outcomes. So if stage 2 would yield outcomes that differ significantly from the Nash/Cournot equilibrium, it seems plausible that this bias would apply to all situations. Since stage 1 of the game is a true one-shot game among the relative attractiveness of the stage 2 outcomes, the obtained results would generally still apply.

3. *Players tend to err on the low side with their prices during the initial rounds of the pricing game.* This result is found in experimental settings.[40] It thus seems relevant to explore the lowest possible price outcomes. Table 4.7 gives the minimax outcomes for the symmetric duopoly explored in chapter 3. The table is derived in the appendix, and shows the lowest possible equilibrium outcomes, according to the Folk theorem. Comparison with table 4.2 shows that the basic structure and outcome of the stage 1 game remain unchanged.

In particular it is easily verified that, like in table 4.2, in equilibrium we get: $\pi_i^{incompatible} < \pi_i^{one-adopts} < \pi_i^{nobody-adopts} = \pi_i^{compatible}$. This means that table 4.7 has again two Nash-equilibria: non-adoption and adoption of compatible versions.

---

[39] See e.g. Tirole (1989).

[40] See e.g. Huck, Normann, et al. (2001) and Plott, Sugiyama, et al. (1993): in small settings, players eventually reach 'collusive price-levels', but with four or more players this appears to be very difficult.

TABLE 4.7 Minimax outcomes for a symmetric duopoly

Firm 2 \ Firm 1	No adopt	Own version	Compatible
No adopt	$\frac{1}{4}$ $\frac{1}{4}$	$\frac{1}{4} - \frac{b-c}{4}$	$\frac{1}{4} - \frac{c}{4}$
Own version	$\frac{1}{4} - \frac{b-c}{4}$		$\frac{1}{4} - \frac{b}{4}$
Compatible	$\frac{1}{4} - \frac{c}{4}$	$\frac{1}{4} - \frac{b}{4}$	$\frac{1}{4}$ $\frac{1}{4}$

4.4.2 Comparison to the model of Shy (2001)

In his book on network industries, Shy applies essentially the same model to a wide variety of industries, including the banking industry. His model is very similar to the model used in this chapter, except that Shy uses switching costs, where my model uses horizontal differentiation and transportation costs.[41] Yet when Shy applies his model to ATMs he finds that firms generally prefer incompatibility. This result is different from my own and from reality. Hence a comparison is in order. I first describe his general model; his application to ATMs will be treated further down.

Shy considers two firms A and B with an equal number of customers, η. These customers obtain services at prices f_A and f_B. Both firms run incompatible networks with benefits equal to $\alpha\eta$. These benefits accrue to customers. Customers start with either of two banks but can switch at a cost δ. Given a price f_B, firm A now can lure (all) customers from B by offering a price of $f_B - \delta + \alpha\eta$: bank A compensates the customers for the switching cost δ, but on the other hand the new customers will now enjoy a network of size $2\alpha\eta$ where as previously they had benefits of only $\alpha\eta$, thus raising the price A can charge. If the networks are compatible, the 'lure away price' for A becomes $f_B - \delta$. Since the price response function is discontinuous (a marginally lower price doubles A's market share) Shy uses the concept of Undercut Proof Equilibrium that I also use in section 4.1.2: both firms set a price low enough, so that the other firm cannot increase profits by undercutting its rivals price and doubling its market share. Shy also introduces an option where one firm opens his network to the other firm unilaterally. Table 4.8 summarizes the resulting equilibria.

[41] And in Shy (2002), the model is used to estimate switching costs using the prices charged by banks, much in the same way as I will estimate transportation costs in next chapter of this thesis.

TABLE 4.8 Summary of Shy's (2001) results

Stage 1 outcome	Equilibrium prices	Equilibrium shares	Equilibrium profits	Welfare
1. Incompatible versions	$2(\delta - \alpha\eta)$	$\frac{1}{2}$	$2\eta(\delta - \alpha\eta)$	$2\alpha\eta^2$
2. Compatible versions	2δ	$\frac{1}{2}$	$2\eta\delta$	$4\alpha\eta^2$
3. Firm A offers compat.				$3\alpha\eta^2$
-firm A	$2\delta - \frac{2\alpha\eta}{3}$	$\frac{1}{2}$	$2\eta(\delta - \frac{\alpha\eta}{3})$	
-firm B	$2\delta - \frac{4\alpha\eta}{3}$	$\frac{1}{2}$	$2\eta(\delta - \frac{2\alpha\eta}{3})$	

TABLE 4.9 Results of Shy's approach using my parameters

Stage 1 outcome	Equilibrium proces	Equilibrium shares	Equilibrium profits
1. Neither adopts	1	$\frac{1}{2}$	$\frac{1}{2}$
2. Incompatible versions	$1 - b + c$	$\frac{1}{2}$	$\frac{1}{2} - \frac{b}{2}$
3. Only one firm adopts			
-adopting firm	$1 + \frac{c}{3}$	$\frac{1}{2}$	$\frac{1}{2} - \frac{c}{3}$
-other firm	$1 - b + \frac{2c}{3}$	$\frac{1}{2}$	$\frac{1}{2} - \frac{b}{2} + \frac{c}{3}$
4. Compatible versions	$1 + c$	$\frac{1}{2}$	$\frac{1}{2}$

It easily follows that consumers prefer (socially suboptimal) incompatibility, while compatibility is the dominant strategy for firms. Shy does not consider the adoption decision, but it is easy to do so by incorporating a cost of adoption per customer. In fact, it is straightforward to make Shy's model compatible with my own. Normalizing the number of customers to 1, Shy's parameter η is equal to my s_i, while his α becomes my b. Finally, Shy's δ translates into half of my transportation costs t; since I normalized $t = 1$, I set Shy's $\delta = \frac{1}{2}$. Using this notation, but otherwise following Shy's approach, I get the equilibria shown in table 4.9.

Comparing these results to table 4.1, the similarities are obvious. In particular the basic structure of the payoff matrices (tables 4.2 and 4.3) remains unchanged and non-adoption is again a (welfare suboptimal) Nash-equilibrium.

The advantage of my own model is that it uses the concept of Nash-equilibrium which cannot be done with Shy's model. In addition, because my market share function is a continuous function of prices, my model can analyze the phenomenon of 'tipping' using the result of DePalma and Leruth (1993).

Finally a word on Shy's ATM model. He changes the network benefit function from $\alpha\eta$ to $\alpha(a_A)$ with a_A being the number of ATMs, which he considers fixed. This changes the nature of the model in a crucial way. If the number of ATMs is fixed, getting more customers *does not* increase the size of the network. This explains his results: the smaller firm likes compatibility, the larger doesn't; and because compatibility requires consent of both, the outcome will be incompatibility, which is inconsistent with reality. The number of ATMs is only exogenous in the short run. If a bank gets more customers, it can place more ATMs, and thus we are back to the original model, where fierce competition destroys profits for both. The result of Matutes and Padilla (1994), that compatibility is not an equilibrium outcome, may be due to the same mechanism. They too consider the number of ATMs as fixed.

4.5 Conclusions

Section 1.5 formulated four questions which are the focus of this thesis. The first two questions concern: (1) the causes of initial differences among countries in payment systems; and (2) the reason these differences persist even if there is agreement that some of the technologies used are suboptimal. Why did some countries adopt ACH/giro technology, while others continue to use the economically inferior check technology?

In chapter 3 I reformulated these questions in terms of equilibrium outcomes and lock-in. Lock-in was defined as a situation where players are in a suboptimal (Nash) equilibrium: no individual player has an incentive to move even though all may be better off in another equilibrium. I then applied a model that assumed standards are unsponsored to analyze equilibrium outcomes. The model showed that various types of lock-in may occur: (1) no player in a (closed) system adopts, (2) semi-autarkic players adopt incompatible versions; and (3) a semi-autarkic subgroup of firms (banks in a country) may adopt while other subgroups fail to do so. Just as importantly I found that (unsponsored) network technologies are 'national': different (incompatible) versions cannot coexist with a group of consumers that transact randomly ($\delta = 1$).

The current chapter analyzed whether this pattern changes if standards are sponsored. The short answer is no, as long as the network effect is not stronger than the existing differentiation between firms ($b < t$):

1. Somewhat counter-intuitively, sponsoring of standards *increases* the range of parameters where of lock-in may occur (as long as network effects are moderate). While sponsoring allows a firm to internalize part of the externality in adopting a network technology, it also enhances the level of price competition, which reduces profits and makes unilateral adoption unattractive. This result holds for all levels of demand elasticity and autarky and for a wide range of industry structures:

   (a) In a symmetric duopoly with perfectly inelastic demand, any single player that unilaterally adopts the network technology (as a first mover) will be worse-off than before, so lock-in can *always* occur. This contrasts with the case of non-proprietary standards, where lock-in can occur *only* if the cost-benefit ratio of the new technology exceeds the share of the largest firm.

   (b) Price sensitive demand mitigates this somewhat; but even for infinitely elastic demand, sponsored standards perform only as well as unsponsored standards in preventing lock-in.

   (c) In more general oligopoly settings, a coalition that jointly adopts the network technology needs a share of more than $\frac{c}{b}$ to make unilateral adoption profitable. Since a share of $\frac{c}{b}$ is good enough under unsponsored standards, this again means that the range of parameters for which lock-in can occur is wider with sponsored standards.

2. 'Autarkic' players or networks (whose customers interact more with each other than with members of other networks) generally prefer their own proprietary standard, since it insulates them from competition. This forms an additional hurdle for migration to a common standard, on top of the migration costs faced in the case of either proprietary or non-proprietary standards.

3. The better performance of unsponsored standards in preventing lock-in comes at a cost: since firms do not compete on network standards, prices and profits are higher, and hence social welfare is lower than with sponsored standards. The importance of this effect rises with ε. For $\varepsilon = 0$ the effect is nil. If $\varepsilon \leq 1$, the adoption of unsponsored standards still realizes more than 90% of the social gains of sponsored standards (if lock-in is avoided).

4. These results hold as long as the network effect is not stronger than the existing differentiation between firms ($b < t$, or actually $b < 1$ since I normalized transportation costs to 1). For large b sponsoring does indeed

decrease the area where the suboptimal equilibria can occur, while for $b >> 1$ sponsoring effectively prevents the occurrence of such equilibria.

Overall, the models show how in a world consisting of semi-autarkic countries with a more or less fragmented banking sector in each country, a new technology can lead to a patchwork of national versions of g and non-adoption by some countries. This can occur if: (1) the network benefits are smaller than existing firm differentiation; and (2) transactions demand is not too price sensitive. Payment technologies appear to meet both these criteria.[42]

[42] The small network benefits will be shown in chapter 6. For price sensitivities, see section 2.1.3.

4.6 Chapter appendix: symbols used in chapters 3 and 4

Parameters:

b : benefits per customer, if all customers use the same version of g

c : costs per customer of network technology g

c_i : coalitions of players: $c_1 = \{1, .., m\}$, $c_2 = \{m+1, .., n\}$,

δ : level of autarky: $\delta = 1$: random tx patterns, $\delta = 0$: autarkic countries

ε : demand elasticity

f : base case technology (not subject to network effects)

F : upgrade of base case technology

g : network technology

k : concentration index; $k \equiv \Sigma_i \frac{\hat{s}_i}{2 - \hat{s}_i}$

n : number of firms

$\hat{s}_i$: natural market share of firm i, defined as share of branches, outlets etc.

$\hat{s}_{c_i}$: natural market share of coalition i, $\hat{s}_{c_i} \equiv \Sigma_{j \epsilon c} \hat{s}_j$

s_l : *share of largest firm or country*

s_c : critical share needed to unilaterally adopt g

t : unit transportation costs (disutility of buying a less than perfect product)

Variables:

p_i : price (net of marginal cost) charged by firm i

$\hat{p}_i$: hedonic price, $\hat{p}_i \equiv p_i + 2 - bs_g$

$\hat{p}_{c_i}$: average hedonic price for coalition: $\hat{p}_{c_i} = \Sigma_{j \epsilon c_i} \hat{s}_j p_j$

$\overline{p}$: average price of $n$oligopolists, weighted by natural shares: $\overline{p} \equiv \Sigma_i \hat{s}_i p_i$

$\overline{\hat{p}}$: average hedonic price of $n$oligopolists, weighted by natural shares: $\overline{\hat{p}} \equiv \Sigma_i \hat{s}_i \hat{p}_i$

s_g : market share of firms that have compatible versions of g.

s_a : share of all countries where banks have adopted g.

s_i : actual market share of firm i

s_{c_i} : actual market share of coalition i, $s_{c_i} \equiv \Sigma_{j \epsilon c_i} s_j$

π_i : profit of firm i

W_C : consumer welfare

W_F : firm welfare (sum of profits)

W_S : social welfare: $W_S = W_C + W_F$

$*$: equilibrium values (as in p^*, s^*, π^*); either Nash- or Undercut Proof-Eq.

5

Case 1: adoption of giro-systems

This chapter analyzes the decision by the Dutch banks to establish a joint giro-system in the mid 1960s, by applying the models of the previous 2 chapters. In particular I estimate the value of crucial parameters (like b, c and s_1) to see whether the model can adequately explain actual behavior by participants. I find that this is indeed the case.

The remainder of this chapter proceeds as follows. Section 5.1 lists the case approach with data sources etc. Section 5.2 describes what actually happened: how giro was introduced in the Netherlands. Section 5.3 describes the economics of giro-systems, to lay a basis for applying the models. Section 5.4 then applies the models of chapter 3 and 4. The last section discusses the results.

5.1 Case background and methodology

The analysis in this chapter rests on: (1) a reconstruction of the actual events; and (2) an estimate of the main variables of the models described in the previous chapters. Actual events were reconstructed using a variety of sources, including some of the literature reviewed in chapter 2, and several books that describe the formation of the Dutch payment system such as Wolf (1983), Peekel and Veluwekamp (1984) and PCGD (1973).[1] In addition, I was able to consult several documents, such as correspondence between the main decision makers, through the archive of S. Lelieveldt.

I used the following information to estimate the model variables:

- The economics of giro payments were determined using the annual reports of the Dutch Postgiro ('Postcheque- en Girodienst') from 1918 to 1969, available through the library of DNB, the Dutch central bank. For each year these contain detailed information on operational costs, revenues, transaction types and volumes, and number of accounts. In addition, I used the results of a detailed study by Flatraaker and Robinson

[1] All of these sources are in Dutch. For an English description see Lelieveldt (2000).

(1995) of the costs of check- and giro-payments in Norway (the Norwegian payment system is very similar to the Dutch system).

- The Dutch banking landscape at the time (1966) was assessed using figures on share of assets, number of branches, and margins using annual reports of the major banks (available through DNB) and the yearly survey of NIBE ('bankenboekje' of Nederlands Instituut voor Bank- en Effectenbedrijf).

- All figures were converted to EUR at 2000 price levels, using Dutch inflation index figures covering 1900-2000 (from the Dutch Statistic Agency CBS), Norwegian inflation figures 1994-2000 (from the Norwegian central bank) and exchange rates from the IMF.[2]

5.2 How giro was introduced

A giro transfer is a way to settle payments. It represents an innovation over the older check technology. A check is essentially an instruction to the bank of the payor (the debtor) to pay the payee (the creditor) a certain amount. The payee then takes the check to his own bank, who presents it to bank of the payee for payment. If there are sufficient funds in the account of the payor, money is transferred to the bank of the payee, who credits the payee's account.

This sounds cumbersome and it is. The process generally involves transporting paper between banks, and if the payor's account has insufficient funds (the check 'bounces'), the whole chain has to be followed backwards to the payor. It is estimated that the US spends well over 1% of GDP on writing and processing checks. Most European countries rely on giro-systems, which are estimated to cost half as much.[3] In a giro-system, the payor instructs his own bank to credit the payee's account, which can be with another bank. His bank transfers money to the bank of the payee, who credits the payee's account.

While simpler than checks, a giro-system requires participants to share common standards; these generally involve a common account numbering system (so the payee's bank and account can be unambiguously identified), a common format for instructions (with e.g. clear rules for accompanying messages and payment information), maintaining settlement accounts at a common institu-

[2] Throughout this thesis, I convert USD to EUR at par. For the Norwegian estimates of Flatraaker and Robinson (1995), I use NKR=0.125 EUR. In addition, I use inflation figures from the Norwegian central bank to convert 1994 prices to 2000 levels by applying a price index of 113 for 2000 (1994=100).

[3] Both the check and the giro cost figures are taken from Humphrey, Pulley, et al. (2000).

TABLE 5.1 Giro-systems in Europe

Country	Year of introduction	Giro transactions (millions, 1958)	Transactions per capita (1958)
Austria	1883	131	18
Belgium	1913	47	5
Denmark	1920	76	17
Finland	1940	33	7
France	1918	706	14
Germany	1908	1027	14
Netherlands	1918	295	25
Italy	1918	n/a	n/a
Luxenbourg	1911	n/a	n/a
Norway	1942	37	10
Sweden	1925	218	29
Switzerland	1906	253	45
Total		2823	16

Source: Thompson (1964)

tion, common rules for finality of payment, exception handling etc. This makes giro-payments a network technology.

In most Western economies giro clearing was introduced in some form during the period 1890-1920. The notable exceptions are the UK and its former colonies (US and the commonwealth). Thompson (1964) gives an overview which is reproduced in table 5.1. Generally this introduction was done through a some form of public initiative. Cooperation between governments, central banks and cities led to postal giro-systems, that used the post offices for access to accounts, and/or by municipal giro-systems. These giro-systems required participants (mostly businesses and public institutions) to maintain giro-accounts. It was only during the expansion of banking into the mass customer segment in the late 1950s and 1960s that banks in most European countries either joined these giro-systems or established their own giro-transfer systems to facilitate transfers between accounts at different banks.

In the Netherlands, the first discussions about a giro-clearing take place in the late 1800s. In 1902-1904, the passing of a new bank law leads to a renewed discussion about the role of the state. For example, should DNB (the central bank) offer accounts to non-banks? The issue is not resolved. Following the debate, various chambers of commerce call for a giro-system using the German model. In 1910, parliament passes a motion to create a giro-service. Following this, the minister of agriculture contacts DNB and the cash-associations to

discuss matters.[4] In April of that same year DNB announces it will have no objections to a joint clearing house of the cash-associations, unless the state takes its own action. The state indeed does this, through the Postgiro act, which leads to the creation of the Postgiro (PCGD) in 1918. This public institution made transfers between customers who had to maintain an account for that purpose. The post offices served as access points to the system. Earlier, in 1916, Amsterdam had already created a municipal giro-system. Several other large cities also established their own giro-system. All of these were absorbed by the Postgiro, although the largest of them, the Amsterdam municipal giro (GGA) did not join until 1979.[5]

In addition, each major Dutch bank had developed its own internal payment transfer system for transfers between customers. Each bank had its own forms and account numbering system, and transfers between customers of different banks were cumbersome. Wolf (1983) describes in detail how the banking sector set up its own system in 1967, after nearly 25 years of deliberations, committees, etc.[6] One of the first reports that describes how such a system for the banks could work was presented in 1943 by Keegstra, a retired head of the Amsterdam municipal giro (GGA). From his own experience he knew what it takes, and his report already mentions the three key ingredients for such a system: (1) a common and central account numbering system, (2) standardized forms and public education, and (3) a clearing house.[7] The banks were not enthusiastic: they saw the extra costs but not the benefits. Instead they opted for streamlining existing procedures for transfers between banks.[8] Over the next 20 years a string of committees studied options for an improved payment system. Finally in May 1965 the Dutch banking association formally told its members "that the board and policy committee had concluded that further detailing of the idea of a central giro clearing house merited further intensive study".[9] To this end they appointed a group of external experts, led by Starreveld, an accountant with KKC (now KPMG). All of a sudden things

[4] These cash-associations facilitated settlement among specific groups, such as the dealers on the stock exchange.

[5] Although the formal merger took place in 1979, the two institutions had for a long time maintained similar forms and account numbering systems. It was quite easy to transfer money between the two institutions.

[6] Wolf (1983), p. 8-49.

[7] Op. cit. p. 21-22.

[8] The Dutch banks were not noted for their progressive views. An (unconfirmed) story has the CEO of ABN making the following remark about why his bank would not adopt ATMs: "People that need money at 3 a.m. cannot have honourable intentions." (mensen die om 3 uur 's nachts geld nodig hebben kunnen geen eerzame bedoelingen hebben); this was in 1983, when the US already had ATMs at every streetcorner.

[9] Op. cit., p. 39.

went quickly: in November 1966 the banks decided to recruit a head for a (still to be founded) central clearing house, and the first computers were bought in March 1967. In July 1967 the commercial and agricultural banks formally founded their own joint giro clearing exchange, called the BankGiro Centrale. As part of this move to a joint giro clearing exchange, all the participating banks adopted common transfer forms, a common account numbering system and rules for handling exceptions, errors and disputes. They also planned (and later executed) campaigns to educate the consumer about how to use these instruments.

Several authors mention two causes for this sudden acceleration of events.[10] In the first place, the commercial banking sector had just completed a major restructuring: in 1964 two mega-mergers of four players with each about 5% share created ABN and AMRO who then had about 10% each (26 years later, in 1990, these two merged to form ABN AMRO). After these mergers the two large players acted as leaders for the commercial banking sector. The commercial banks represented 33% of the total banking market, and they cooperated closely with the agricultural banks, which had another 28% of the market. In the second place, by the mid-1960s, several large banks decided to make a major move into the mass market; Postgiro had already a significant presence in this segment because many government employees received their salary through a giro account. The commercial and agricultural banks decided that they needed a high volume efficient payment system to execute the high transaction volumes brought by mass banking.

Interestingly the banks did not join the Postgiro-system. Wolf (1983) describes the events that led to the decision to go it alone. By the mid-1960s banks were losing payment share to Postgiro, and they were concerned about this competition by a public institution. Early plans for a clearing house included procedures where the banks would check if the beneficiary of a transfer had a bank account, even if the instruction mentioned his Postgiro account. If that were the case, they would credit the bank account, thus keeping the money in their system.[11] The atmosphere was therefore not exactly cooperative to begin with. The actual rift appears to have been caused by the technicalities of the systems. Postgiro was planning to shift to transfer forms that were in effect punch cards. The banks wanted to stick to their carbon paper instruction forms. In addition, the banks wanted to use a 10 digit numbering system with a check digit; Postgiro used 7 digit numbers without a check digit,

[10] E.g. Wolf (1983), p.39.

[11] In all fairness Postgiro did the same, they would only transfer money to another giro account, forcing banks to maintain giro accounts to receive payments. This requirement was imposed on the Postgiro by law.

but was unwilling to assign all its customers a new account number. Knowing that the Postgiro had committed itself irreversibly to punch card forms, the banks still stuck to their choice for flexible paper forms, thus confirming the incompatibility of the two systems.[12]

Afterwards there was a bit of a blame game going on about who should have conceded to whom. Wolf (1983) states that the Postgiro was not very cooperative.[13] A key figure of Postgiro, Reinoud, would later claim Postgiro *did* offer cooperation, while the banks were not open to such cooperation.[14] In any case, both parties were unable to agree, thus creating two separate and somewhat incompatible systems.[15] For the next two years after 1967, both camps courted the joint savings banks (who held 15% of the market). In 1969 the savings sector decided to join the system of the commercial banks. The two systems remained separate until the 1990s. Under increasing pressure from the Dutch National Bank the Postgiro and the banks gradually implemented a set of technical and procedural changes effectively creating one system by 1997.[16]

5.3 The economics of giro-systems

Exact data on the giro-systems are hard to obtain, since the majority of the costs are made by banks, where they tend to be mixed up with other costs. I therefore use data on the Dutch Postgiro from the 1920's, and 1950/1960's. This has a double advantage: it provides data on a player whose only business was maintaining accounts for giro clearing, and these data cover the period when the relevant choices were made. Figure 5.1 shows the cost of a giro-transaction during these years.[17]

Some interesting observations can be made. There were significant economies of scale in the early years: costs per transaction declined rapidly as volume grew. After a failed attempt to switch to Hollerith technology in 1923, the

[12] Wolf (1983), p. 33.

[13] "PCGD niet toeschietelijk", op. cit., p. 41.

[14] Reinoud of Postgiro stated in an interview: "In 1966 prof. Starreveld got the assignment to set up a Bank Clearing House. At the time I discussed with him whether it would not be optimal to set up one large technical centre. Both banks and Postgiro would then be able to use this, while keeping their identity. The time was apparently not ripe for this idea. I wonder however, whether the National Clearing System would not have been realized in an easier way, if we would have pursued this plan at the time." (Hogesteeger and de Lanoy Meijer, 1992, p. 55-56, my own translation from the original Dutch text).

[15] The incompatibility was not total. For transfers across the two systems one did not have to revert to checks. It was cumbersome though. If a Postgiro customer wanted to transfer money to a bank account he had to instruct the Postgiro to transfer the money to the giro account held by that particular branch of the bank, adding the actual bankaccount number

Figure 5.1 Cost of a giro transaction in EUR at the 2000 price level
Source: PCGD annual reports.

system remained essentially manual, until 1962.[18] This manual system was surprisingly cheap: in the period 1928-1934 it performed on average 52 million transactions per year, at an average cost of EUR 26 million per year (converted to current value), or EUR 0.50 per transaction, much less than the EUR 1.34 that a paper giro cost in 1994 in Norway.[19] Remarkably, the large scale automation (1962-1964) did not lead to lower costs per transaction; it did however stop the increase caused by the rise in real wages after the war.

---

to be credited as an extra instruction. In addition there were delays: a transfer across systems took 3-5 days, a transfer within a system took only 1-2 days.

[16] Press release Dutch central bank: www.dnb.nl/persberichten/1997/nbc.htm

[17] The Post giro transaction figures include cash withdrawals, deposits and a few checks. The majority of transactions were giro transfers: they represented 79% of all transactions in 1959 (the last year for which the annual reports give a breakdown by transaction type). The gaps in the graph correspond to the failed automation move in 1923-1924 and WWII, including the postwar reconstruction years.

[18] The story of this failed switch is quite colorful. The failure was so complete, that the Postgiro had to shut down operations for a full year to sort out the mess. It even had to ask customers to provide proof of their account balances, because the internal administration could no longer be relied upon. An official inquiry found many causes, among them an overly optimistic external management consultant (!), who had a contract entitling him to a quarter of all labor savings during the first year (Tak and Dubois, 1924).

[19] Norwegian data from Flatraaker and Robinson, 1995, converted to EUR year 2000 prices. Several explanations can be offered for this seemingly paradoxical result. First real labor costs more than quadrupled in the period 1928-1994. Second, the functionality of a 1994 giro payment was much richer than in 1928: printed statements, added information etc.

TABLE 5.2 Key data on costs and output for the Dutch giro-system, 1918-1934.

	Costs (mln EUR)	Transactions (millions)	Accounts (thousands)
1918	2.6	1.3	13.7
1919	5.2	2.8	21.8
1920	8.9	4.6	32.6
1921	13.6	7.6	55.4
1925	16.4	18.4	113.2
1926	17.0	23.4	120.0
1927	17.3	27.9	129.3
1928	19.4	31.9	139.3
1929	19.1	37.1	151.3
1930	22.7	46.1	167.5
1931	26.2	52.7	184.8
1932	28.3	57.8	204.3
1933	31.5	64.4	224.3
1934	31.9	71.1	245.0

Source: PCGD annual reports. Costs are converted to 2000 EUR level.

Table 5.2 gives the key data for the pre-war period.[20] Regressing total real costs ($COST$) on the number of accounts (ACC) and transactions (TX) for the period 1918-1934 for the Dutch giro bank, yields:

$$COST = \underset{(t=2.6)}{3.5} \text{ million} + \underset{(t=3.0)}{110}\, ACC + \underset{(t=0.3)}{0.34}\, TX;\ \text{Adj. } R^2{=}96\%,\ DW{=}1.24$$

This suggests a fixed costs of running the system of EUR 3-4 million per year, and a per account cost of around EUR 100 per year per account. The costs do not appear to significantly depend on the number transactions. This last observation is important, since this is exactly the cost structure assumed by the models in previous chapters.[21] These models do not assume fixed costs, however the fixed costs in the Dutch giro-system were relatively small, they represented about 10% of the total costs in 1936, when the giro had almost

[20] I use these years since in this period the number of accounts and number of transactions behaved somewhat independent. For later years the number of transactions per account was almost constant (at about 400), creating multicollinearity between the two variables.

[21] It is also somewhat improbable: not *all* costs are driven by maintaining accounts. At least part of the costs have to be directly related to the transactions themselves. However, for the early giro-system a large part of the cost may indeed have been account related: all customer account balances were maintained in ledgers, with each balance being updated daily, by hand. Since some doubt lingers about the true cost structure (which part is account related and which part is transaction driven) I will test the outcomes of the analysis against various cost structures.

300,000 accounts. Of course as the number of accounts grew, the relative importance of fixed costs declined even further.

5.4 Applying the model

Using these data, I now examine the decision by Dutch banks to establish their own giro-system in 1967.[22] I first estimate the key parameters of the model of chapter 3 (unsponsored standards) to calculate the critical share given by proposition 3.1 in that chapter. I find that the adopting banks had indeed a joint share above this critical level. Since there are indications that both camps (Postgiro and banks) had the option of excluding the other (the standards were de facto sponsored), I also estimate the variable t of the sponsored model, to see if parties had indeed incentives to maintain incompatible systems.

5.4.1 Estimating parameters for the unsponsored model

Proposition 3.1 states that a group of banks will only unilaterally adopt a new network technology if $s > c/b$, where s is the joint share of these banks, c is the fixed per customer cost of allowing him to use the new transaction technology and b is the benefit to the bank if a customer uses the new technology for all his transactions. Using 1966 data from the Netherlands, I estimate c and b to derive the minimum share that is needed to profitably adopt the technology. The first column of table 5.3 summarizes the estimates. They were derived as follows.

In 1966 (the year the decision for a bankgiro-system was made) total costs of the Postgiro were EUR 254 million, for which they performed 393 million transactions on 1.3 million accounts (these and subsequent figures have all been converted to 2000 price levels). This implies 309 transactions per account and an integral cost of EUR 195 per account. Based on the earlier regression results, I assume that all costs are driven by the number of accounts and not by the number of transactions (further on, I will test the sensitivity of results to this assumption). I further assume fixed costs to be negligible for the purpose of this exercise, given the earlier modest estimates of fixed costs (EUR 3.5 million, or 10% of 1936 cost and 1.5% of 1966 costs). Finally I assume that the total costs of the Postgiro system reflect *additional* cost on top of a check system. Thus I get c = EUR 195, where c is the fixed cost per customer per year of the new technology.

[22] Table 5.1 confirms that the Netherlands is a good representative of the more successful giro countries.

TABLE 5.3 Critical share for giro adoption in the Netherlands: sensitivity of results to alternative assumptions for cost structure

	100% of giro costs and 0% of check costs are account driven	Alternative 1: 50% of giro costs are account driven	Alternative 2: 25% of check costs are account driven
Fixed cost/account (EUR)	195	98	195
Cost/giro transactn (EUR)	0.00	0.32	0.00
Cost/check (EUR)	1.42	1.42	1.07
Transactions/person per yr.	309	309	309
Parameter c	195	98	195
Parameter b	439	340	330
Critical share c/b	44%	29%	59%

I also need an estimate for the extra benefit per transaction of using the new technology, parameter b of the model. To get the value of b, I put the cost of a check in 1966 at EUR 1.42. I get this figure by applying a multiple of 2.2 (based on both the Norwegian and the US experience) on the integral cost of a Dutch giro transfer of EUR 0.64.[23] Now $b = (1.42 - 0)$ x $309 = 439$; multiplication with the number of transactions per year is necessary since the model normalized the number of transactions per person to one.

Using proposition 3.1 of the previous chapter, the critical share is then equal to $c/b = 195/439 = 44\%$. Thus one or more players need a joint market share of at least 44% to be able to unilaterally and profitably adopt the new giro technology.

Therefore the Dutch commercial banks (who had only 33% of the market in 1966) did indeed need all of their members and the cooperation of the agricultural banks (28%) to make the economics work. No individual bank had a market share even close to what was needed to go it alone.[24]

There are two critical assumptions underlying this result. The first critical assumption is of course the cost structure, and in particular the fact that all giro costs are driven by number of accounts, while none of the check costs are.

[23] I get the EUR 0.64 using the 1966 Postgiro data(converted to year 2000 prices). The reason for multiplying this by 2.2 instead of directly using the 1995 Norwegian estimate of check costs is the following. The Norwegian data on a mail giro gives a cost of EUR 0.92, much higher than the Dutch figure of EUR 0.64. I suspect this reflects differences in methodology. Therefore it seems safe to take the factor difference between check and giro cost that was obtained for Norway, 2.2, and apply it to the Dutch estimate for the cost of a giro transaction.

[24] In 1918 the Postgiro was established with a much lower share. However, this was a public initiative and it may be argued that this institution was to a large extent autarkic, which significantly reduces the necessary critical market share.

To test for the robustness of the estimates against changes in the cost structure, table 5.3 gives the results for two alternative sets of assumptions. If only 50% of all giro costs are driven by number of accounts (and the remainder of giro costs is transaction related) the critical share decreases to 29%. If only 75% of all check costs are transaction driven, the critical market share increases to 59% (if combined, the two alternative assumptions cancel out: the critical share is then 43%).

The second critical assumption is that standards are unsponsored, or that firms at least do not use the technology as a competitive weapon (by passing part of the benefits to the consumer). As discussed at the end of section 5.2 it is not entirely clear whether parties saw the technology as something to be shared or not. The outcome suggests they were able to keep the system proprietary to some extent. Therefore I now apply the model of chapter 4 to see what would have been rational behavior if the standard was indeed proprietary.

5.4.2 Estimating parameters for the sponsored model

To apply the sponsored standards model, I use the parameters b and c derived above. In addition I need estimates for customer and firm differentiation (parameter t), natural market shares ($\hat{s}_i$), and price sensitivity (α). I use observed ('real world') equilibrium prices to estimate transportation costs t as follows. In the standard 2 firm Hotelling equilibrium we have in equilibrium $p^* = t$, where p^* represents prices net of marginal costs. The model for asymmetric oligopolies in chapter 4 is a bit more complicated, but the principle is the same. Equation (4.6) expresses equilibrium prices relative to transportation costs t. To estimate t we thus need parameters $\hat{s}_i$ and p_i^*. I estimate $\hat{s}_i$ by the share of branches of each bank in the total number of branches in the country. These are given in table 5.4. p_i^* is estimated from $\hat{s}_i$ by applying equation (4.6) from section 4.2.[25]

If we assume that equilibrium prices are indeed determined according to the model in chapter 4, then the model leads to $p^* = 1.31t$ for the largest bank, and $p^* = 1.21t$ for the average bank. For the Dutch banking sector, I estimate p^*_{ABN} to be in the range EUR 807 to 3063.[26] This implies $t = p^*_{ABN}/1.14 =$

[25] This p^* is normalized for $t = 1$ so in reality it should be interpreted as a multiple of t.

[26] Based on profit figures for 1967 for the Dutch banking sector, converted to 2000 price levels. Since the estimate of b =EUR 439 is based on a period when Postgiro had mostly business customers, we need an estimate for p that represents a similar mix. I have therefor taken 1967 figures for market leader ABN. They had 346.000 accounts in 1967, a large part of these were businesses. Revenues per account were EUR 3063, while profit (before tax) per account was EUR 807; Since I have no insight into their cost structure at the time, I take the

TABLE 5.4 Share of branches for Dutch banks, 1966

Bank	Branches	$\hat{s}_i(\%)$	p_i^*	$\frac{\hat{s}_i}{2-\hat{s}_i}$
Girobank	2526	31.4	1.31	0.19
Raifaissen	1364	16.9	1.21	0.09
Boerenleenbank	1036	12.9	1.18	0.07
Savings union	800	9.9	1.16	0.05
AMRO	695	8.6	1.16	0.05
ABN	469	5.8	1.14	0.03
NMB	368	4.6	1.13	0.02
NCB	99	1.2	1.11	0.01
Savings bank Amst.	92	1.1	1.11	0.01
Other	599	7.4	1.11	0.04
Total	8048	100	1.21	0.54

Source: NIBE bankenboekje (1967)

EUR 708 to 2686.[27] The normalized value of b is then $439/t$ =0.16 to 0.62, well below the critical level of $b = 1$.

Finally, I need an estimate of price sensitivity. A thorough econometric estimate of price elasticities for payment instruments is given in Humphrey, Kim, et al. (2001). Using data for Norway covering the period 1989-1994, they find elasticities of 0.50, 1.07 and 0.29 for ATM withdrawals, checks and EFTPOS respectively. I use the average of these: $\varepsilon = 0.62$. Using equation (4.5) this corresponds to a price sensitivity of $\alpha = 0.21$.[28] In summary, I get the following

full range as an estimate for p^*_{ABN} (remember that p represents the mark-up over marginal costs).

[27] This is somewhat higher than the switching costs reported in Shy (2002) for the Finnish banking system. Given the similarities between his model and my own (see section 4.4.2) his switching costs and my transportation costs are measuring more or less the same thing, where Shy's switching cost δ correspond to half my transportation costs. Shy finds switching costs of \$ 400-464, using lifetime discounted fees. Since my transportation costs are annual and his switching costs are 'lifetime' the difference is very substantial. Two explanations are (1) Shy looks only at listed retail consumer fees, where my approach incorporates all revenues (fees, interest margin etc.) across all customers (wholesale and retail); (2) Shy looks at retail customers, while my estimates are based on figures that include mostly corporate customers (in 1967 ABN had few private customers).

[28] Rewriting expression (4.5) I get:

$$\alpha = \frac{\varepsilon}{p(1+\varepsilon)+2-\varepsilon}.$$

Taking $p = 1$ and $\varepsilon = 0.62$, this gives $\alpha = 0.207$. The expression (4.5) was derived for a symmetric duopoly; I assume it holds more or less for the oligopoly model.

normalized values:

$$\begin{aligned} b &= 0.16 - 0.62 \\ \frac{c}{b} &= 44\% \\ \alpha &= 0.21. \end{aligned}$$

5.4.3 *Applying the model for sponsored standards*

With these values we can now study the adoption of a sponsored giro-technology by Dutch banks. From figure 4.3 I conclude that for the values of $\alpha = 0.21$ and $\frac{c}{b} = 44\%$ lock-in can indeed occur if the market structure is a duopoly. Figure 4.2 indicates that this conclusion holds for values of b up to about 2.2, which is reassuring giving the uncertainty surrounding the estimate of t.

The market structure in the 1960s was, however, not a duopoly. Before the mergers of 1964 it resembled the polar case of 'Gorilla vs. competitive fringe' with the Postgiro as a Gorilla with a 31% share of payments.[29] After the mergers the landscape shifted a little towards the polar case of 'Gorilla vs. fringe', with $n = 5$ to 10 (depending on how one treats the savings and cooperative banks). For these situations figures 4.4 (Gorilla vs. fringe) and perhaps 4.5 (equal sized players) would apply. The figures were drawn for $\alpha = 0$ and $n = 1000$, while we have $\alpha = 0.21$ and $n = 5$ to 10; both of these changes shift the curves in the graphs to the right.

This would put the situation before the mergers ($\hat{s}_1 = 31\%, \frac{c}{b} = 44\%$) in the dark shaded area of figure 4.4, where lock-in and compatibility are the equilibria. However, if the curves shift to the right (because $\alpha > 0$) and $\frac{c}{b}$ is lower (e.g. because of a different cost structure) we have compatibility as the only equilibrium, *if the fringe is able to act jointly.* If the fringe is unable to do so, the Gorilla can improve profit through unilateral adoption, while the fringe is then unable to form a counter coalition. This leads to semi-adoption as the only equilibrium. And indeed this was the situation from 1918 to 1967. After the 1964 mergers, the fringe was able to organize itself. And it indeed adopted its own version of the technology. It would then have been rational for parties to establish compatibility. This is more or less what happened, although it took almost 30 years and a lot of pressure from the central bank.

[29] Postgiro was a payment specialist. In the more general banking market it was hardly a Gorilla.

5.5 Discussion of results

Overall, the model appears to do a plausible job of explaining the emergence of giro clearing in the Netherlands. Public action created the first giro-system in the early 1900s. By the 1960s this system had a solid position against a more fragmented banking sector. As the banking sector become more concentrated, they were able to take joint action, and also adopt the giro technology. And, eventually, parties settled on the equilibrium of one compatible version of that technology. More generally, most countries that adopted giro-systems in the early years of the 20th century now use ACH/giro rather than checks. Where giro-systems did not exist (US, Canada and UK) banks stuck with the check system; the one exception to the rule is France (who else), which had an active giro-system in the 1950s and 1960s, but currently still uses checks on a large scale. Obviously, closer analysis of these other countries is needed before we can apply the "Dutch explanation" elsewhere. Rather than offer a deterministic explanation for country differences, I would like to make a more general point: analysis of the Dutch events demonstrates that the adoption of payment systems can indeed be characterized by the two properties of network externalities: path-dependence and excess inertia.

The model does *not* explain why it took the sector 30 years to establish compatibility. As described earlier, it is not even clear who refused compatibility to whom. One explanation for this could be non-random transaction patterns, i.e. $\delta < 1$. Thus far I assumed transaction patterns were random across institutions. While this is true now, it was certainly not true in the early 1900s. In fact, many of the municipal giro-systems were viable only because transactions occurred disproportionately within one city. One could speculate that by 1967 traffic still took place disproportionately within rather than across the two sectors (Postgiro and banks); for example all public entities effected their payments through the Postgiro, while most businesses used banks. According to proposition 4.3 and figure 4.1 this could have created an area where it was in the interest of both firms to maintain incompatibility. Since this lowers social welfare, it also explains why both parties were not keen to admit this publicly. Following this reasoning, the advent of mass consumer banking could have randomized traffic patterns: most transactions now take place between businesses and the public sector on one side and consumers on the other, rather than between businesses and between public entities. This could have pushed the situation from the lower band of figure 4.1 to the middle part, where both players prefer compatibility.

Finally, the question why some other countries did not adopt ACH/giro remains to be answered. As was argued before, the absence of the chance event

of early giro introduction may have played a role. For the US lack of market concentration probably played an important role as well. Table 6.3 in the next chapter gives an overview of the market concentration in various countries. The concentration of the US banking sector is lower than in any other country in the table: the top 10 banks represent less than half the market. In 1990, before the major consolidation of the 1990s, the top 10 banks held just 20% of deposits, while in the 1960s (when the Dutch banks adopted their clearing house) it was even less.

Table 6.3 also shows that the Canadian and UK banking markets are very concentrated: the top 3 institutions hold more than half the market; and indeed transfer use is on the rise in these countries. However, in terms of overall instrument use they are somewhat in between the US and continental Europe. This suggests that concentration alone cannot explain everything. Other factors, like the absence of giro-systems, must also have played a role.[30]

[30]The UK did adopt a giro-system, but much later than the other countries.

6
Case 2: European harmonization of ACH systems

Most European countries have gone through similar developments as Holland, and by the introduction of the physical Euro in January 2002, almost all participants had a giro clearing system. While these systems enable large volumes of national payments to be cleared efficiently, cross-border transfers remain both difficult and expensive. Following increasing pressure from the European Commission, the European banking sector held a workshop in 2002 to discuss options to harmonize their systems.

This chapter applies the models of chapter 3 and 4 to analyze the situation and the actions taken by banks from an economic point of view: what were the likely motives for the participating banks and can the model help to understand and explain the outcomes of the workshop? I find the answer to be affirmative.

Section 6.1 describes the main sources of information used in the case. 6.2 describes the actual events. Section 6.3 derives estimates for the main parameters of the models, after which section 6.4 uses the model to analyze the motives of the various stakeholders. The last section compares the actual outcomes with the model predictions and discusses the results.

6.1 Case background and method

To reconstruct actual events I used documents from the EU (directives, press releases, discussion papers and research reports), from the European Central Bank (ECB), and from the banking sector itself, notably the European Banking Federation (in particular the May 2002 white paper on SEPA) and the European Payments Council (EPC, which was installed following the SEPA workshop). I also discussed the SEPA workshop and its results with two participants: G. Hartsink of ABN AMRO and R. Heisterborg of ING.

The estimates of the model parameters were made using the following data sources:

- Transaction patterns across European countries were analyzed using data on SWIFT messages between each country. These data are available though BIS. To analyze cross border card transactions, I used (public)

data from VISA and MasterCard. Analysis of these patterns served as a basis for an estimate of the level of autarky across countries (parameter δ).

- Market shares of the major players (parameter r_{ij}) in each European market were calculated using balance sheet data from annual reports and (for some countries) national central banks. The share of each country in Europe (parameter s_i) was based on population data available through BIS.

- Finally, for the economics of giro technology (parameters b and c) I used the estimates derived in the previous chapter.

6.2 Creating a Single Euro Payments Area (SEPA): description of events

In 2002, the introduction of the physical Euro created a single payment instrument that could be used throughout 12 EU countries. While there was now a single cash instrument, the same could not be said for other instruments. Several countries had their own electronic purse scheme, a Belgian consumer could not use his debit card at most Dutch merchants and while ATMs throughout Europe accepted practically all cards, it was generally more expensive to use a card abroad than at home. Perhaps the biggest problem was formed by cross-border transfer payments. While transfers within countries generally were free and fast, cross-border payments remained expensive, slow and prone to errors.

6.2.1 1990s: "Europe" grows increasingly frustrated with cross-border transfers

By 2002, the European Commission had been urging banks to improve this situation for several years, as it considered the lack of a proper Pan-European payments infrastructure an impediment to the further integration of Europe. A 2000 note from the commission to the European council and parliament radiates frustration with the banks' lack of progress: "The cost and performance of cross-border credit transfers has been a concern for the Commission since many years. Already in 1990, the Commission gave an in-depth analysis of the problem in 1990 (COM(90)447 - 'Payments in the European Single Market') emphasizing that in the 90s structures were established to ensure payment services between Member States which are as inexpensive, quick and reliable as domestic systems. In 1992, the Commission established a work programme

(SEC(92)621 - 'Easier cross-border payments: breaking down the barriers'), indicating that there was a need to improve retail cross-border payment services before the completion of the EMU. In 1994, COM(94)436 - 'Fund transfers in the EU: transparency, performance and stability' was published which included the proposal for a Directive on cross-border credit transfers."[1] This directive was indeed adopted in 1997. It covered payments up to EUR 50,000 and laid out targets in three areas:[2]

- Transparency: banks were obliged to inform both parties in a transaction of any fees and commissions, including how these were calculated. The sending party has to be able to specify whether these charges are to be borne by himself ("OUR-transfer"), the beneficiary ("BEN-transfer") or shared with the beneficiary ("SHARED-transfer").

- Minimum standards regarding execution times (6 business days) and distribution of charges ("no double charging"), well as a money-back guarantee in case a transfer gets lost (a not uncommon event).

- Complaints and redress schemes.

Banks were given 30 months to comply with these minimum standards. A 2001 survey conducted for the European Commission sent out 1,480 credit transfers of EUR 100 using 40 bank accounts in 15 member states. The survey found progress on execution times: 2.97 days on average, compared with 4.61 and 4.79 days in earlier surveys conducted in 1994 and 1993 respectively. Reliability was found to remain an issue, with 1% of transfers not arriving (two-thirds of these were returned to sender, one-third went missing in the system: the sender's account was debited, but the beneficiaries account was never credited). The most serious findings, however, concerned the fees and commissions. First, it proved quite difficult to avoid charges to the beneficiary: while all transfers were sent as OUR-transfers, the beneficiary was still charged in 16.2% of all cases. Moreover, charges appeared to have gone *up* over the previous years. The study found that a cross-border transfer of EUR 100 cost an average of EUR 17.36, including charges to both the sender and the receiver; instead of declining, charges had increased by 1.55% compared to a similar survey in 1999.[3]

---

[1] Quoted from "Communication from the commission to the council and the European parliament", 2000. Available at http://europa.eu.int, document nr 5200DC0036, accessed on Sept. 23rd 2003.

[2] Directive 97/5/EC (European Commission, 1997).

[3] European Commission (2001).

6.2.2 2001: legislation on pricing of Euro payments

Basing itself on this lack of progress, the European Parliament passed regulation EC 2560/2001. This regulation forced banks to maintain the same tariff structure for domestic and international Euro payments below EUR 12,500, and to implement a common account numbering system (IBAN).[4] Compared to earlier directives, these were drastic measures. In most countries domestic transfer payments are free or priced below EUR 1 per transaction, so banks would lose almost all revenues on cross-border transfers.[5] To give an impression of the impact on profits: in 2002 banks in the seven EU countries covered by BIS executed about 70 million cross-border transfers. Assuming 75% of these were for amounts below EUR 12,500, banks faced a potential loss of close to one billion euro per year. Without a decrease in processing cost, this reduction in revenue would go straight to the bottom line.[6] Furthermore, the directive went beyond transfer payments (the topic of the earlier EC directive and the 2001 study) and targeted card payments as well; card payments comprise 80% of cross-border payments, and the economic impact for especially the banks in the Northern countries was substantial: their customers use their cards to withdraw money from ATMs in tourist destinations. Banks charged the card holder a fee in the order of EUR 1.50-3.00 per cross-border ATM withdrawal to recover the inter-bank fees charged by the owners of these ATMs. Since in many countries (e.g. Netherlands, Belgium and Austria) domestic ATM withdrawals were free, the harmonization of prices meant a substantial loss of income. The Dutch banks for example, faced the loss of more than EUR 50 million in card revenues.[7] Since costs were not affected, this represented an equivalent reduction in profits.

Finally, the directive set tight deadlines: for cards the regulation was effective as of July 1st 2002, and for transfers a year later. The banking sector was clearly caught off-guard: several bankers expressed that they had not expected such harsh measures.[8]

---

[4] Regulation no. 2560/2001 (European Community, 2001).

[5] This assumes banks would price cross-border euro transfers as domestic transfers. The alternative would have been to raise prices on domestic transfers, but this has proved to be impossible in most countries.

[6] The calculation assumes they would lose revenues of EUR 17 per transaction on 75% of 70 million transactions=893 million Euro.

[7] The estimated loss is based on an estimated 25 million cross border ATM withdrawals for which the banks charged an average of 2.25 Euro. The alternative would have been to charge domestic ATM withdrawals as well. So far the banks have been unable (or unwilling) to do this.

[8] For the Dutch reader: one of the driving forces behind the regulation was Karla Peijs, then member of the European parliament; she became the Dutch minister for transportation in 2003.

6.2.3 2002: the SEPA workshop

The EC directive drastically increased the urgency for banks to take action. There was an existing Coordination Group on European Payment Systems (COGEPS). Chaired by G. Hartsink of ABN AMRO, it consisted of representatives of several large EU banks, the European Central Bank (ECB), and the 3 European Credit Sector Associations (ECSAs).[9] This group organized a joint workshop of 40 banks and the 3 ECSAs, which was held on 25-26 March 2002.[10] The purpose of the workshop was to discuss joint actions on harmonizing the national payment infrastructure, thus creating a Single Euro Payments Area (SEPA). Three main areas were covered: (1) transfer payments; (2) card payments; and (3) governance of the implementation of any joint actions. Following the workshop the participants issued a white paper (European Banking Federation, 2002b), containing the main decisions. I describe the main outcomes below.

Transfer payments

The outcomes on transfer payments roughly follow the main three areas needed for a joint system: a joint clearing house, a common account numbering system and common forms, rules procedures etc. With regard to a joint clearing house the banks considered 5 models:

1. A single clearing house that would handle all Euro transfers, both domestic and cross-border.

2. A pan-European clearing overlay, where a new clearing house would handle the Euro cross-border transfers, while the domestic transfers would continue to be handled by the domestic clearing systems.

3. Direct linkages between the existing national clearing houses. Banks would not send their Euro cross-border transfers to a joint clearing house. Instead they would send them to the national clearing house, after which the national clearing houses would exchange them.

4. Bilateral exchanges between individual banks (this would mean continuing the existing correspondent banking system).

5. Using the card network of VISA and/or MasterCard.

---

[9] There are three of these ECSAs: one for the commercial banks, one for the savings banks and one for the cooperative banks.

[10] European Banking Federation, 2002a.

The participants selected option 2 (a pan-European overlay). Their arguments were as follows. Option 1 (a single clearing house) was seen as involving disproportionate migration costs: all domestic systems would have to change, clients would have to be issued new account numbers etc. The banks did however envision that some of the small countries (with a subscale and expensive local clearing system) would move domestic volume to the pan-European system. At some point in the future ever larger countries might do the same, thus gradually realizing a single European system in the long run. Option 3 (direct links between the national clearing houses) was seen as impractical, because it would introduce an extra step in the process. In addition the national clearing houses were thought to be insufficiently capitalized to handle the systemic risk arising from cross-border transactions.[11] However, the participants left a form of option 3 open as a way for smaller banks to access the pan-European clearing house. Option 4 (bilateral exchanges between banks) was rejected because it would imply a continuation of the existing situation. Option 5 (use a card network) was considered very interesting for person to person transfers, but less so for corporate and SME transfers; there were concerns about counterparty risk if the networks (designed for retail transactions) would be used for larger transactions.

With regard to a common account numbering system, banks agreed that migrating to a common system was too costly. Instead they chose to use the International Bank Account Number (IBAN) for cross-border transfers. This IBAN is in effect an overlay over the existing national systems, adding a country and bank code to existing national account numbers (it is described in more detail in section 6.3). The main problem with this IBAN is a misalignment of incentives: the bank of the *sending party* has to supply the IBAN of the receiving party with the payment instruction. However, if the IBAN is missing or incorrect, the bank of *receiving party* bears the cost of (manual) rework. Therefore banks decided to either (1) introduce an inter-bank tariff that differentiates between STP (Straight Through Processing, which can be done if all information is correct) and non-STP payments; or (2) start rejecting non-STP payments altogether by 2005.

Finally, all banks agreed that the lack of common rules, forms and procedures was a major hurdle. However, much of these rules are set by national legislators, not by the banking sector. It is almost impossible to develop a pan-European

[11] A logical question is why this was not a problem at the national level. The reason is that at the national level, the ACHs only process information, the actual net payments are made through the national central banks; the counter party risk is therefore an exposure against the national bank. This does not work at the European level because each bank still maintains his account with the local central bank, not with the ECB.

direct debit product, for example, because each country offers different legal protection and recourse in case of disputed debits. The task of coming up with common rules and procedures was delegated to the European Committee for Banking Standards (ECBS) and the national governments.

Card payments

For cards, there is already a pan-European system, in fact even a global system, that works reasonably smoothly. For credit cards, several players offer solutions that are truly global in terms of technical infrastructure, product features and pricing. For PIN-debit cards the situation is a bit different.[12] For ATMs acceptance is truly global: most debit cards will work in any non-proprietary ATM across the globe.[13] Cross-border acceptance of PIN debit at the Point of Sale is less general, but PIN debit cards can increasingly be used at foreign stores.

Cross-border debit card transactions at both ATMs and POS are generally processed through an 'overlay', such as Maestro (operated by MasterCard). One option is to migrate domestic debit card systems to the common overlay. The main barrier to doing this is formed by substantial differences between the Euro countries in inter-bank pricing of POS transactions. In the Northern countries this interchange is either zero or small, while in the southern countries the issuing bank receives a substantial amount per transaction from the acquiring bank (comparable to credit card transactions). A single system would presume a common pricing structure across Europe; migrating to such a common system will thus always cause substantial migration pain in some countries.

As a result, the workshop led to no real concrete steps on card payments, other than to study the issues and possible solutions.

Governance

The workshop led to the formation of the European Payments Council (EPC), roughly consisting of the workshop participants. It is headed by a steering committee, consisting of 10-15 senior bankers. The actual work is carried out by several working groups that periodically report to the steering committee. This sounds like the United Nations, and to some extend it is. The EPC has little power to enforce measures, and the group represents a fragmented banking industry with interests that diverge between large and small banks (represented

[12] Signature debit cards are MasterCard and Visa cards where each transaction is directly deducted from a current account. They are just as global as the credit cards of those brands.

[13] Banks in some countries have proprietary ATMs that can only used by their own customers.

through the ECSAs), northern and southern banks, and consumer and corporate banks. Its main instruments are therefore persuasion, and involving senior executives of major institutions who can use their clout in their own country to get the other banks to move.

6.2.4 Other initiatives

In addition to the SEPA workshop several other initiatives were developed to reach a European infrastructure for transfer payments. Two of these merit mention.

Right before the workshop, VISA (the credit card network) publicly announced a solution enabling person-to person (P2P) transfers using its card network. As said before, this option was not selected at the SEPA workshop. VISA itself has continued to further work out this option, which it is currently offering to banks. It enables cross-border P2P transfers, for example by workers from developing countries who want to send money home.

The national clearing houses, led by the Interpay (the Dutch clearing house) held a series of meetings to explore linking these national systems and use them for international payments. Their shareholders (the banks) somewhat scaled back their ambition by not selecting this as their main mechanism, but such linkages are being studied as a way to give the smaller banks access to a common clearing house.

6.3 European ACH/giro landscape

6.3.1 Incompatible systems

All European countries have at least some ACH infrastructure. This may be one single entity that clears all payments between banks, after which net amounts are settled through the national bank (e.g. France, Belgium, Netherlands). It may also be several systems that each process payments between certain banks and exchange remaining transactions among them (Germany) or it may even be a system of several clearing banks, that exchange payments bilaterally, with the smaller banks accessing the system through one of the clearing banks (UK). However, in all these cases there are national standards for the account numbering system and for the rules and regulations surrounding these payments.

While standards in these areas help to facilitate national transactions, they hinder cross-border transactions. For example, in the Dutch system an account number is a unique identifier and payment instruction forms do not ask

TABLE 6.1 European domestic account number formats: number of digits

Country	Total	Bank Code	Branch Code	Accnt Nr	Chk Digit	Example
Belgium	12	3		7	2	539-0075470-34
France	23	5	5	11	2	20041 01005 0500013M26 06
Germany	18	8	*	10	(1)	53201300 37040044
Italy	23	5	5	12	1	X 05428 11101 000000123456
NL (bank)	10			10	(1)	054 30 26 841
NL (giro)	4-7			max 7		4178233
Sweden	11	4	*	6	1	5491 000000 3
Switzerland	4-21	max 5		max 16	(1)	762 1162-3852.957
UK (dom.)	13-14	6	*	8		60-16-13 31926819
UK (Chps)	15/19	8/11		8		NWBK BG21 01Z 31926819

Note: table shows number of digits, both numbers and letters. Check digit between parenthesis means it is included in account number. * under branch code means branch code is included in bankcode. Taken from ECBS (2003), p. 6 and 7.

for many details of the bank of the recipient. In most other countries however, the account number is not enough and additional information on the bank is needed. The format for such information differs by country (including differences in address system such as different ZIP-code formats). As an illustration, table 6.1 gives an overview of the account numbering systems in the eight European countries that are also part of the BIS-11. In practice, many cross-border transactions require manual rework. Similarly, the lack of standard rules leads to confusion over charges (paid by sender, beneficiary, or both) while recourse in case of lost or disputed payments is often unclear.

To deal with the differences in account numbering systems, the International Bank Account Number (IBAN) system has been developed.[14] Because the account numbering system of each country has a different structure, IBAN (International Bank Account Number) is an overlay that adds a country and bank code to the existing numbers. As a result it has variable length, while it lacks a rigid architecture. It is comparable to the EDIFACT standard; that too was an overlay over existing EDI standards, and for precisely the same reason: the local (national and sectorial) EDI standards were incompatible but too well entrenched to replace.[15]

[14] See ECBS (2003) for a description.

[15] See David and Foray (1994) and Graham, Spinardi, et al. (1995), for a description of this very interesting case.

TABLE 6.2 Foreign transactions as a share of all transactions by country

	pop. (s_i)	Transfers	Cards	Both
Belgium	3%	0.6%	5.9%	2.4%
France	16	0.3	2.0	1.2
Germany	22	0.1	4.9	1.1
Italy	15	0.4	2.8	1.3
Netherlands	4	0.3	4.6	1.7
Sweden	2	0.4	3.3	1.5
Switzerland	2	0.9	3.7	1.8
UK	16	0.4	2.2	1.6
Average	11	0.3	3.0	1.9

Note: the data on cross-border transfers are based on BIS data for domestic transactions and category 1 Swift messages. Each international transfer is assumed to generate 2 such messages. Data for international card payments based on EFIMA estimates for 1997.

6.3.2 Autarkic countries

Assume a world where all consumers transact randomly with each other. For each country i, the share of international payments q_i (as a percentage of all payment transactions in that country) would then be equal to $1 - s_i$, where s_i is the share of that country in the world. Take for example the largest EU country, Germany. It has a 22% share of the EU population. If transaction patterns were random across the EU, one would expect that 22% of all the transactions of a German are with other Germans, leaving 78% for transactions with non-German Europeans.[16] As table 6.2 shows, however, the share of foreign payments is in reality much lower.[17] Instead of 78%, only 0.14% of all German transfer payments are with non-Germans. Transaction patterns for cards are a little less autarkic, but still only 6.3%, not 78%, of German card transactions are made abroad. Note that in line with our model, the share of cross-border transactions is higher for small countries, and lower for larger countries. Also, the share is much higher for card payments than it is for transfers.

For the eight countries in the table, I get an average share of foreign transactions of $\overline{q} = 1.9$; this is an average of card transactions and transfers. For

[16] This of course ignores the world outside the EU. Of all European cross-border traffic in 1999, 84.7% of transactions were between European countries, the remaining 15.3% were with the rest of the world (source, BCG, 2002, p. 43). In the context of my analysis I can therefore indeed more or less ignore the rest of the world.

[17] The eight countries in the table are the European countries that are also part of the standard BIS reports, meaning uniform data are available for them. Jointly, they represent 79% of the EU population.

cards alone, the average share of foreign transactions is higher: $\overline{q}^{cards} = 3.0\%$, while for transfers it is much lower: $\overline{q}^{transfers} = 0.3\%$. I now estimate δ using the equation $q_i = \delta(1 - s_i)$ which was derived in section 4.2.1. Here q_i is the share of foreign payments, and s_i is the overall share in the system of country i. Regressing the data for the seven countries in the table yields:

$$q_i = \underset{(t=14.3)}{0.018}\,(1 - s_i), \quad \text{adj. } R^2 = 27\%.$$

Thus I get an average δ of 1.8%.[18] Using the same approach I get $\delta^{cards} = 4.1\%$, while for transfers alone $\delta^{transfers} = 0.5\%$. Since $\delta = 1$ corresponds to no autarky and $\delta = 0$ represents total autarky, the results imply that the payments world is still very much a national affair. I lack figures on foreign card payments for Canada and the US, but for transfers the share of foreign transactions is 0.05% (US) and 0.2% (Canada) suggesting similar or even larger autarky for these countries.

6.3.3 *Industry structure: locally concentrated, fragmented at European level*

Most European banking markets are relatively concentrated. Table 6.3 gives an overview of the joint market share of the top 3, 5 and 10 banks for each of the major markets. The last column shows the Herfindahl index: $H = \sum_i x_i^2$.[19] At the national level, the top 3 banks represent more than half the market in all five small countries, as well as in two large countries. The banking sector in the other three countries (France, Italy and Germany) is less concentrated, but even there the top 10 banks still form more than half the market. This is reflected in the Herfindahl indices: H is 1000 or higher in the 6 concentrated countries, but even in the three others it is still in the range 100 to 1000.[20] By

[18] The low R^2 is partly caused by the restriction of zero intercept. Without this restriction the regression yields:

$$q_i = \underset{(t=1.66)}{-\ 0.02} + \underset{(2.96)}{0.040}(1 - s_i), \quad \text{adj. } R^2 = 53\%.$$

This result is interesting: is suggests that the relationship between size and autarky is *more* than proportional: large countries are more than proportionately autarkic, small countries are more than proportionately open.

[19] I have included Spain because it is a large European country even though it is not covered in the BIS red books. Jointly the 9 countries represent 90% of the population of the EU plus Switzerland.

[20] To give some perspective on the Herfindahl index: the US justice department considers a market with an H over 1800 to be highly concentrated, $1000 \leq H \leq 1800$ is considered moderately concentrated, while a market with an H below 1000 is considered unconcentrated (FRBSF, 2002).

TABLE 6.3 Bank concentration in major markets, 2000

	Top 3	Top 5	Top 10	Herfindahl index
Netherlands	80%	86	93	2220
Belgium	71	85	90	1920
Switzerland	69	74	79	2430
Sweden	58	74	79	1300
Spain	52	62	72	1140
UK	51	70	89	1190
France	42	60	79	880
Italy	31	42	52	430
Germany	26	36	52	300
Total Europe	8	13	21	50
Canada	60	93	99	1800
US	23	31	44	246

Note: table shows share of top 3, 5 and 10 banks in assets of all banks in each country. Share of top 3, 5 and 10 banks in Europe are obtained by taking share of these banks in assets of all banks in the 9 European countries in the table. US and Canadian figures are 2002 share of deposits. The Herfindahl index is calculated using the largest 10 institutions in each country. Following industry tradition it is expressed on a scale from 0 to 10,000. Source: annual reports.

contrast, at the European level the landscape is much more fragmented: the top 3 banks have just 8%, while the top 10 have 21%. The Herfindahl index for Europe as a whole is 50, or 6 times as low as for Germany, the most fragmented country.

For comparison purposes I have included Canada and US. The Canadian market is comparable to highly concentrated European countries like Belgium and the Netherlands. The US has undergone a major consolidation in the past 20 years, and the concentration of the US banking sector is now at the same level as Germany. Only 12 years ago however, the top 10 US banks had only 20% of the market, comparable to where Europe as a whole is now.[21]

The institutional framework is similarly fragmented at the European level.[22] The banks are organized in 3 European Credit Sector Associations (ECSAs): the European Banking Federation (for the commercial banks), the European Savings Banks Group and the European Association of Cooperative Banks. Most large European banks belong to the first. However, the European Bank-

[21] Figure taken from Rhoades (2000), p. 26. As pointed out earlier, this fragmentation of US banking may help explain its trouble in adopting ACH-transfers on a wide scale.

[22] The importance of institutions in shaping and reinforcing path dependency has been noted by others, such as David (1994).

ing Federation is not a federation of banks but a federation of national banking associations; the banks in turn are members of these national banking organizations. In practice these ECSAs serve as lobbying platforms, not as platforms for joint decision making on payment systems.

A second candidate for cooperation could be payment networks. In most continental European countries banks tend to cooperate at the national level in matters of payment networks; this may in part reflect (continental) Europe's relative tolerance for industry cooperation. In each country, banks typically run 1 or 2 networks for ATMs and or Debit Cards. Table 6.4 gives an overview of the situation in 1988 and 2001 for the 8 European BIS countries plus US and Canada. Note that for almost all countries the number of networks has declined, perhaps reflecting a realization that compatibility is to be preferred (in line with the outcomes of the models in chapters 3 and 4). One option for cooperation at the European level, would be through negotiation between these national networks. Assuming banks in any country act jointly at the country level, the migration to a common standard could take place by negotiation between these national platforms. This would create a more concentrated setting: the top 5 countries represent 78% of the EU plus Switzerland (in terms of population). The problem is that banks in several countries do not act jointly. This is especially true in Germany. In this largest EU member state, there is a clear split between 4 large commercial banks (Deutsche, Commerz, Dresdner and BHV) on the one hand and the savings and cooperative banks on the other. The commercial banks serve large corporations and handle most of the cross border transfers while the savings and cooperative banks serve the vast majority of consumers, and generate most of the cross border card payments. In payments they have often conflicting interests, and as a result the German banks seldom act in unison in payment matters.

If the national payment networks cannot serve as a platform for European cooperation, how about (nascent) pan-European payment networks? There are several candidates: Europay, VISA, Swift and the European Banking Association (EBA). Europay became part of MasterCard in 2002, and as it becomes more integrated in that global organization it becomes less suitable as platform for European cooperation. The same holds for VISA.[23] Swift is an interesting case. The messages that accompany all correspondent banking transactions are it's core business. Swift may have seen a pan-European transfer system as a threat to its message volume. Whatever the reason, Swift did not serve as a platform in harmonizing systems or developing a common solution. EBA is a

[23] VISA did offer to develop a solution for cross-border payments, using its network. But it announced the initiative in a statement that was rather antagonistic to the European banks.

TABLE 6.4 Number of payment networks by country (minimum of ATM and POS)

	1988	2002
Belgium	4	1
Canada	4	n/a
France	1	1
Germany	4	4
Italy	1	1
Netherlands	2	1
Sweden	2	1
Switzerland	2	2
UK	3	1
US	37	14

Source: BIS (1993 and 2003).

private club of European banks. It already runs a system for the clearing of large payments (STEP1), and to reduce counter party exposure, membership is restricted to the larger institutions. As things stand at the writing of this thesis, EBA is indeed becoming the main solution for pan-European transfers, through its STEP2 system. This is a clearing system for small payments between members, where the net settlement takes place through STEP1.

6.4 Applying the model

6.4.1 Model for unsponsored standards

The model for unsponsored standards gives the conditions under which lock-in into multiple incompatible versions of a network technology can occur. I use expression (3.4); lock-in can occur if:

$$\delta s_1 < \frac{c_m}{b}, \text{ for all } i. \tag{6.1}$$

Here s_1 is the share of the largest country, c_m is the per period cost of migrating one customer to the new standard, while b is the per period benefit of using the network technology. The expression assumes all countries in country decide jointly on migration.[24] Because we are looking at ACH systems, I use $\delta = 0.4\%$,

[24] Note that this does not mean the banks in the largest country have to act jointly; in fact they don't have to do anything since presumably the others will move to their standard. If banks do not decide jointly, the expression (6.1) becomes

$$\delta s_1 - (1 - r_{i1})[1 - \delta(1 - s_i)] < \frac{c_m}{b}, \text{ for all } i,$$

the average autarky for transfers (derived in section 6.3.2). Since Germany is the largest country, $s_1 = 22\%$. Thus (6.1) becomes:

$$\delta s_1 = 0.09\% < \frac{c_m}{b}.$$

Using the estimate of b=439 (derived in the previous chapter), we get that lock-in can occur if:

$$c_m > b \text{ x } 0.09\% = \text{EUR } 0.42.$$

In other words: if the per period migration costs are more than half a Euro per customer, a country has no incentive to move to the standard of the largest player. Assuming migration costs are depreciated over 10 years, we get a hurdle of around EUR 5; if one-time migration costs per customer are higher than this hurdle, countries will not unilaterally migrate to the standard of the largest country. What if all banks in all countries would act jointly? The condition then becomes:

$$\delta = 0.4\% < \frac{c_m}{b} \Leftrightarrow c_m > b \text{ x } 0.4\% = \text{EUR } 1.75.$$

Thus if one time migration costs are below EUR 17.50, it would be beneficial to adopt a common standard if all firms in all countries would act jointly. Migration would include changing the account numbering system, issuing new transfer forms, and thoroughly redesigning each bank's payment systems. Therefore migration costs could easily exceed the EUR 17.50 level. Thus one expects banks to opt for an overlay instead of migrating their domestic systems to a common standard.

6.4.2 Model for sponsored standards

How would this change if the national standards are considered sponsored? In that case proposition 4.3 and figure 4.1 suggest that banks may very well like the insulation against competition that incompatibility provides. The very low levels of δ puts us at the left side of figure 4.1, where firms generally prefer incompatibility.[25] If anything, this would give players in a country an extra incentive to resist the migration to a common transfer system.

---

where r_{i1} is the share of the largest player in country i. For our values of δ (~ 0.02) the LHS becomes negative, meaning that even without migration cost, a single player in a country will not deviate from the national standard to adopt the standard of another country. This is fairly obvious: the loss on the transactions with the other players in his country is far bigger than the gain on the cross-border traffic with the country whose standard that firm would adopt, given the very low value of δ.

[25] This holds as long as b is below $\frac{1}{2}$. For larger b the system 'tips' to the standard of either country.

6.5 Discussion of results

As actual events demonstrate, it took rather extreme pressure to get the European banks to improve the (lack of a) system for cross-border transfer payments. And when they did take action, they opted for an overlay system; they did not migrate their domestic systems to a common standard.

The application of the models of chapter 3 and 4 enhances the insight into the blockages faced by the European banks in harmonizing their ACH transfer systems. These blockages are indeed formidable:

- Given the very high autarky of payment patterns, the economic incentive for adopting a joint system is not very strong. One could argue that with a better system there would be more cross-border transfers, but even if cross-border volume double or tripled, it is not obvious that migrating to a common system makes economic sense (indeed one may wonder if the migration to a common currency made sense from a pure transaction point of view; its effectiveness in imposing budget discipline on member states is a different, and debatable, matter).

- Common systems across countries may bring fiercer competition as banks can more easily offer their retail services in other countries. This in turn could depress bank profits, further decreasing their appetite for a common system.

- The industry is very fragmented. The largest 10 European banks have a joint share of only 21%, and there are no strong industry platforms or suitable payment networks to serve as a platform for cooperation.

Given these obstacles, actual events are in line with the predictions of the model. Banks found it hard to establish joint action, and when they did take action, they opted for an overlay (EBA Step2) instead of a common system for all transactions.

It is also quite understandable that the EU government took drastic measures. Even an overlay will facilitate cross-border payments and thus likely increase social welfare. The adoption of a common system would probable further increase consumer welfare through fiercer cross-border competition. While this is clearly good for consumers it is not certain that it would increase overall social welfare: the migration costs, which are borne by banks, may be higher that the gain in social welfare due to more intense competition and the use of a common standard.

7

Technical change and technology succession: introduction and theory

Chapter 3 explored how a country can get locked into an economically inferior (payment) technology by failing to adopt a newer technology; this can occur if the newer technology is subject to increasing returns. It was found that such lock-in may persist even as other countries adopt a better (network) technology if transaction patterns are semi-autarkic. Chapter 4 looked at the role of competition and found that with sponsored standards lock-in may still occur, unless the proprietary network externalities are very large compared to existing firm differentiation. Taken together this reasoning explains why (1) countries tend to act as a whole in the adoption of (payment) technologies, and (2) lock-in persists.

The previous chapters do not consider pre-existing differences in the installed base of payment technologies. The experience in payment systems however, suggests that such pre-existing differences in the technology base play an important role. The description of the US and Netherlands experience in section 1.4 showed how the two countries follow different trajectories; the installed base of payment technologies plays a crucial role in enabling the adoption of certain new technologies. Thus the initial differences transgress the arrival of new technologies "visiting the iniquity of the fathers upon the children and upon the children's children, unto the third and to the fourth generation".[1] As several authors have observed (and as will be confirmed in this chapter) things are not always that gloomy and often the adoption failure of the fathers leads to prosperity for later generations. The old-testamentical damnation is thus replaced by new-testamentical salvation: "but many that are first shall be last, and the last shall be first".[2]

The theory on network externalities explains (albeit at a high level of abstraction) how historical accidents can cause a system to reach a suboptimal equilibrium and adopt a socially inferior standard. Excess inertia can cause lock-in into this standard, preventing the adoption of better technologies. However, it is not clear why countries would follow different paths. Why do differences in the installed base lead to differences in the adoption of new techniques, even

[1] Exodus 34:7.
[2] St. matthew 19:30.

if these new technologies become available to all countries? I therefore look at the broader theory of technical change and technology succession.

7.1 Innovation and technical change

Ruttan (1997) gives an overview of the literature on this topic, and distinguishes three schools of thought:

1. The neoclassical approach of induced growth. Factor prices and latent needs drive an economy to discovering and adopting technologies that satisfy these latent needs and optimize factor inputs given their prices. Important early contributors are Griliches and Schmookler (e.g. 1963).

2. The evolutionary approach, which takes its inspiration from Darwin and Schumpeter (Survival of the fittest, Neue Kombinationen, Creative Destruction), and stresses dynamics and bounded rationality, as opposed to the full rationality and static equilibria of the neoclassical approach. Important contributors are Nelson and Winter (1982), who make extensive use of simulation models to study behavior of firms with different behavior.

3. The path dependence approach of Arthur (1989) and David (1985). Their simpler models allow for more analysis (compared to the evolutionary approach), but that simplicity also limits the applicability. And as Foray (1997) and Cowan and Foray (2000) show, path dependence is easy to prove in theory, but finding empirical evidence is very difficult.

The book of Dosi, Freeman, et al. (1988) bundles contributions from the latter two approaches. The contributors often see themselves as "heretics" against the "Newtonian and Cartesian" orthodox economic theory.[3] As the editors write in their introduction: "This book emerged out of growing dissatisfaction felt by a number of economists and non-economists alike with the way technical change has been and continues to be treated in mainstream economics." Where the neoclassical approach analyzes equilibrium and assumes perfect knowledge of agents, the other two schools of thought produce disequilibrium models with bounded rationality of agents. Dosi, Freeman, et al. were not the first to revolt. Schmookler himself, in his 1966 book, questions the causality in the relationship between R&D and technological progress. Analyzing time-series data on growth and patents by industry, he finds that patents follow growth, and not

[3] Dosi, Freeman, et al. (1988), p. 199, p. 409.

the other way around. He suggests that the simplest explanation may well be the correct one: industries with more value added have more money to spend on R&D, irrespective of the potential for further technological progress. With this he reverses the conclusion from his earlier work with Griliches. In their 1963 article they report that inventive activity is distributed along the value added of sectors, and suggest that this confirms the neoclassical maximizing behavior.

In another break with the neoclassical approach, Atkinson and Stiglitz (1969) challenged the notion that technological change shifts the entire neoclassical production function, and instead introduced localized technical change.[4] The authors point out that this leads to technology paths. The concept of technology paths is taken further by Dosi (1982). He proposes to define paradigms as cylinders in some multi-dimensional space. Normal technological progress (intra-paradigm) follows a path within such a cylinder, driven by the evolutionary forces of Schumpeter and the 'natural trajectories' of Nelson and Winter. Selection among paradigms (inter-paradigm progress) is driven by economic and institutional forces and a trend towards labor saving, and thus more induced/neoclassical. In a later publication, Dosi (1997) explicitly links the second and third stream, arguing that they are highly complementary. Firms act as a conduit and selection mechanism for new technologies according to the model of Nelson and Winter, much like individual animals and plants act as a selection mechanism for genes. The system as a whole then follows the technology paths with the increasing returns and path dependence of Arthur and David.

Freeman and Perez (1988) propose a taxonomy of technological changes and distinguish four types: incremental changes, drastic changes, new systems and paradigm shifts. This last type of change would correspond to the paradigms of Dosi, and according to Freeman and Perez it occurs when a key input to the economy: (1) experiences a drastic drop in cost; (2) is in ample supply; and (3) has the potential for use in many applications. They relate each of five Kondratieff cycles to such a change in a key input: cotton and pig iron (1770s), coal and transport (1820s), steel (1880s), energy and oil (1930s) and micro-electronics (1980s).

---

[4]The word 'local' can have two meanings in this context. Atkinson and Stiglitz use local to describe a specific point on the production function, as opposed to a geographic locality. As they themselves point out there is a link: different economies may start their improvements from different points in the production function (e.g. due to different factor endowments), and this may indeed to geographically local technical change: the advances of the West may not be relevant for developing countries.

A similar distinction is made by Breshanan and Trajtenberg (1995), in their study of General Purpose Technologies (GPTs), like the steam engine, the electric motor and semi-conductors. When applied to the different sectors, these GPTs produce technological progress, often in the form of more incremental improvements.

Several authors make explicit mention of the analogy with the concept of punctuated equilibrium, taken from biology.[5] New technologies are proposed to follow a product life cycle: in the early stages there are many variations, but early on a dominant design emerges, which is then incrementally developed further. Kauffman, Lobo, et al. (2000) show how this pattern may arise naturally: if the current technology gives poor results, it pays to perform (costly) searches far away from the current practice. However, once an adequate technology (or process) is found, it is better to search nearby; searching farther away gives results closer to the mean, i.e. below the practice already found previously. Abernathy and Utterback (1975) use data from five US industries to support their hypothesis that industries go through three phases, affecting both the production process and commercial strategy. Gort and Klepper (1982) find similar patterns using data from 46 industries. They find that most of these pass through five stages: (1) a single firm introduces a radically new product; (2) a sharp increase in the number of producers; (3) exiting firms start to balance new entrants; (4) shakeout where new entry is negative; and (5) stability. Gort and Klepper show that this pattern cannot be explained by scale economies.

Many authors focus explicitly on the role of firm competition in new technology introduction. Tushman and Anderson (1986) suggest that radically new technologies are often competence destroying and thus shape the industry as new firms (with the required new competencies) take over. They support their analysis with data on firms in 4 industries. However, several other authors point out that new technologies do not always destroy existing players. For example, Tripsas (1997) analyzes the typesetting industry, and finds that only one of the three new technology waves led to a restructuring of the industry. Similarly Rothaermel (2000) shows how symbiotic cooperation between start-ups and incumbents in the biotech industry appears to benefit both.

It is interesting to note that the changes in the payments industry have not (yet) led to a Schumpeterian gale of destruction across the banking industry, and change in the use of payment instruments does not appear to be driven by innovative firms destroying incumbents. However, as will be shown in chapter 9, the industry did appear to pass through the stages of Gort and Klepper in

[5] E.g. Nelson (1998), p. 329.

the late 1990s when the promise of Internet payments lured a flood of new entrants; however almost all of these failed, while the incumbents survived.

In summary, much of the literature embraces the idea that technical change follows trajectories and is path dependent. The mechanism by new technologies spread through an industry and an economy has been modelled and analyzed empirically. However, this literature does not explain why countries that have access to the same technologies still follow different paths. The next section therefore looks at the literature on country differences in technology.

7.2 Technology and country differences

Structural economical differences between countries have long fascinated economists. As Prescott (1998) convincingly argues, a factor of 40 difference in labor productivity between the West and some developing countries cannot be plausibly caused by differences in capital stock (using the neoclassical production function). However, neither can they be fully explained by differences in technology: after all, access to technological know-how is supposedly free. Arguably, there is a time-lag as the low productivity countries adopt the new technologies, so one would expect these countries to catch-up with the advanced countries, with the rate of convergence slowing as they approach the leaders. Abramovitz (1986) describes this process of catching-up. However, he immediately points to two problems: (1) it is not supported by empirical evidence: the process of convergence showed marked strength only during the first quarter-century following WWII; and (2) it cannot explain why leading countries fall behind.

Several of such leadership changes have been described. Veblen (1915) describes the economic rise of Germany before WWII (surpassing the UK in areas like chemicals).[6] Another high-profile leadership change was the Japanese success in (car) manufacturing in the 1980s.

Patel and Pavitt (1998) review persisting differences in number of patents filed, R&D spending, education and training between OECD countries for the period 1970-1990; they find the differences to be significant and structural. They go on to hypothesize different innovation systems, attributing the post-war success of Germany and Japan their 'dynamic' innovation system as opposed to the UK and US 'myopic' innovation system. The success of Japan in particular has been a rich source of explanations for country differences. Japan's success has been attributed to such diverse factors as a critical domestic consumer base (Porter, 1990), a more dynamic innovation system (Patel

[6] In this respect Veblen (1915, p 126) talks of "penalties of taking the lead": a newer nation outperforms the incumbent.

and Pavitt, 1998), and a superior Asian work ethic (see Gong and Jang, 1998, for a review and clear rejection of this theory).[7] Recent history has not been kind to these theories: the current lack of economic performance of both Japan and Germany effectively kills the theory of Patel and Pavitt about Japan's (and Germany's) supposedly 'dynamic' innovation system as opposed to the 'myopic' innovation system of the US and UK.[8]

A more promising approach is taken by Nelson and Wright (1992) in their analysis of the "Rise and fall of American technological leadership". They attribute the changes in position to the national nature of technology: a lot of knowledge is complex, tacit and implicit, advances are often incremental and local (as in Atkinson and Stiglitz, 1969), and technological progress is network phenomenon, requiring interaction of many firms and people and subject to (local) increasing returns. As a result technology is subject to path-dependence and lock-in at the national level. A similar argument is offered by Lundvall (1988); he describes technological progress as a highly interactive process, subject to local increasing returns. Keller (2002) performs an econometric analysis on technology differences between countries and finds them related to geographical distance: the amount of knowledge spillovers is halved every 1,200 kilometers; these results appear to confirm the hypothesis of Lundvall. Keller also finds that language skills are important, as there appears to be more technology sharing between the English speaking countries, while overall technology knowledge appears to become more global over time.

Jaffe and Trajtenberg (1999) study patent citations from US, UK, France, Germany and Japan, to analyze international knowledge flows and spillovers. They find that patents whose inventors reside in the same country are 30 to 80% more likely to cite each other than inventors from different countries. In an earlier publication, Jaffe, Trajtenberg, et al. (1993) found similar 'autarky' within the US: patents from the same state or metropolitan area were found to be more likely to cite each other.

The effect of such local technology externalities (spillovers) has been modelled by Krugman (1994), Dalle (1997) and David, Foray, et al. (1998). Krug-

[7] There is, as Gong and Jang note, much "confusion on Confucianism" and its economic role. While Max Weber (who explained the success of the West through its protestant work ethic) in the 1950s found Confucianism to be a cause of Asian backwardness, by the mid-1980s several authors explained the Asian economic miracle by that same Confucianism. And, to complete the circle, in 1998 Francis Fukuyama regarded Confucianism as the root of the Asian crisis (story and all sources taken from Gong and Jang, 1998).

[8] That innovation systems can be a tricky concept is perhaps illustrated by the conflicting diagnosis of European/German industry. Where Patel and Pavitt (1998) praise Germany's 'dynamic' innovation system, Amable and Boyer (1995) blame Europe's (and by implication Germany's) 'linear' innovation system for the fact that Europe is falling behind the US.

man finds that local externalities do indeed lead to country specialization (with multiple equilibria, i.e. the specialization of a country depends on historical accidents rather than natural endowments). Dalle and David, Foray, et al. describe closely related models with agents on a lattice selecting one of two technologies subject to local externalities. The authors find that local spillovers lead to either coexistence of technologies each in their own region (small global externalities) or pure standardization on one technology (strong externalities). These models and outcomes are closely related to those of Cowan and Cowan (1998) described in chapter 2. Indeed, if one postulates that technological systems are subject to local increasing returns, all the models of paragraph 2.2.3 can be applied.

And that is precisely the problem with these models: they circumvent the issue of technology succession, by stating that technology follows a path of incremental change (within a 'paradigm') that is subject to increasing returns. However, they neither operationalize the notion of incremental change (is debit cards replacing guaranteed checks incremental?) nor do they explain if (or why) paradigm shifts would occur according to local patterns. To explain the paths in payment technologies, the issue of technology succession will have to be addressed explicitly.

7.3 Models of technology succession

Very few models of technology succession exist. Almost all theories focus on the *new* technology: how it is found, refined, adopted and diffused; very little is explicitly said about the technology being replaced. A notable exception is Shy (1996), who models succession of various versions of a network technology in an overlapping generations model; each period a new and improved version of the technology arrives, but incompatibility with the installed base leads to substantial switching costs for existing agents. Over time, the gap between the installed base and the 'state-of-the-art' grows, and new generations (not committed to the installed technology) arrive on the market. At some point the system switches to the state-of-the-art technology, and so on. Windrum and Birchenhall (2000) point out that Shy's model is a good start, but in its extreme simplicity it fails to take into account many factors that play a role in technology succession; after reviewing the literature on technology succession the authors conclude that little if any modeling work has been done on succession (without contributing a model themselves).

In summary, the theory reviewed in this chapter has a lot to say about the process of adoption and features such as path dependence. As such it helps explain the phenomenon of specific country paths. But it does not offer any clues as to how this path dependence behaves. The next chapter introduces a model of technology succession to fill this gap.

8

A model for succession of payment networks

This chapter describes a model of technology succession that explicitly takes into account the role of the installed base in the decision to adopt new technologies. In does so by assuming that technologies share certain cost and benefit components. If a new technology shares one or more cost components with technologies in the installed base, its adoption will become more attractive, because those cost elements have already been provided for. Conversely, its adoption will become less attractive if it shares benefit components with technologies in the installed base, because those benefits are already provided by existing technologies.

After describing the model in some detail, I use the model to analyze three questions:

1. Does path dependence occur, i.e. can differences in adoption choices lead to different equilibrium outcomes? and if so, how prevalent is this phenomenon?
2. What is the economic significance of such path dependence: by how much does it influence overall welfare?
3. What factors influence the occurrence and economic significance of path dependence?

I find that path dependence can occur even in the simple '2 by 2' case of 2 technologies in an environment with 2 cost and benefit components. The number of different paths rises with the number of technologies and components. The economic impact of following different paths can be substantial.

Does this mean that countries are indeed likely to adopt different technologies and end up with different welfare levels? Yes: I find that small differences in the initial starting position indeed lead to the adoption of different technologies in a significant share of all cases. Such differences lead to significant differences in welfare across countries. Interestingly, this phenomenon is hardly reduced by global scale economies that reward the adoption of technologies that are used by other countries. Does timing matter? Again the answer is yes. Even if two countries both start with a clean sheet (or an identical technology base), differences in the timing of adoption of subsequent technologies can

still cause technological divergence. Countries that delay the initial adoption of technologies are generally better off, because they can pick from a wider array of technologies, while the country that started earlier with adoption may be hampered by an installed base of early (inferior) technologies.

Finally, I use the model to generate patterns of technological progress. I find a preference for small incremental changes, but larger jumps also occur. This suggests that 'paradigm shifts' are just extremes of a continuous distribution of technological change, and thus generated by the same process that also produces incremental changes.

The remainder of this chapter proceeds as follows. Section 8.1 describes the model, section 8.2 examines the occurrence and economic significance of path dependence. Section 8.3 applies the model to analyze technology paths across countries, and section 8.4 explores what the model has to say about patterns of technological progress. The last section summarizes the findings.

8.1 Description of the model

8.1.1 Basic elements

The model distinguishes two levels. At the higher level there are technologies. In the context of payment instruments, two examples of such technologies could be Point Of Sale debit (POS) and Automated Teller Machines (ATM). At the lower level, each technology uses a number of cost components to deliver a number of benefits. An example of a cost component of POS could be the merchant terminal. For ATM it could be the teller machines.[1] The same approach is used for benefits. These too come in components. Using the payment example, for ATM, the benefit component could be the labor saving on human tellers at the branch, while for POS the benefit component could be that same labor saving on human tellers, plus the consumer convenience of being able to spend at a store, without first having to get cash. The two technologies thus share a benefit component.[2] Table 8.1 gives an overview of the cost and benefit components used by ATM and POS technologies.

[1] Obviously both technologies need a card in the consumers wallet, and this is a cost component they share. For the example I have assumed away the cost of these cards, to keep the number of cost components at two (the card and the merchant terminal). The example thus fits the 2 by 2 case (2 technologies and 2 components for both costs and benefits) that will be analyzed in the next section.

[2] The example is illustrative and highly simplified. In practice, ATMs provide benefits that debit cards cannot provide, such as the ability to get cash for person-to-person payments, or for paying at stores that do not accept debit cards. As said, the example is purely illustrative.

TABLE 8.1 Benefit and cost components of ATM and POS technologies

	Benefits		Costs	
	Teller savings	Convenience	Teller machines	Terminals
ATM	1	0	1	0
POS	1	1	0	1

Formally, I distinguish (a list of symbols used in this chapter is given in the appendix to this chapter):

Benefit components	$i = 1..m$
Cost components	$j = 1..m$
Technologies	$k = 1..n$.

For each technology k, the relevant benefit and cost components can be thought of as two (transposed) vectors $\mathbf{b}^k$ and $\mathbf{c}^k$. In the example we would get for ATM: $\mathbf{b}^{ATM}$={1,0} and $\mathbf{c}^{ATM}$ ={1,0}; and for POS: $\mathbf{b}^{POS}$={1,1} and $\mathbf{c}^{POS}$ ={0,1}. These transposed vectors are taken from table 8.1. Technologies are adopted if the benefits exceed the cost. To determine these, the benefit components i and cost components j have to be multiplied by a value. Following the approach in the previous chapters, these values could be expressed as benefits or cost per customer per period for a particular component. To keep the model simple I assume all cost components have a cost per period of 1, while all benefit elements have a value per period of 1.1; thus a technology with an equal number of cost and benefit elements will be slightly profitable. In our example we would get:

$$\begin{aligned} \pi_{ATM} &= 1.1 \sum_{i=1}^{m} \mathbf{b}_i^{ATM} - \sum_{j=1}^{m} \mathbf{c}_j^{ATM} \\ &= 1.1 \sum_{i=1}^{m} \{1,0\} - \sum_{j=1}^{m} \{1,0\} = 0.1 \end{aligned}$$

and

$$\begin{aligned} \pi_{POS} &= 1.1 \sum_{i=1}^{m} \mathbf{b}_i^{POS} - \sum_{j=1}^{m} \mathbf{c}_j^{POS} \\ &= 1.1 \sum \{1,1\} - \sum \{0,1\} = 1.2. \end{aligned}$$

Here π represents overall social profit. If the benefits and costs accrue to firms, this profit accrues to firms, otherwise π may be shared by firms and consumers.

In any case π corresponds to the difference between benefits and costs, $b - c$ in the notation of earlier chapters.

We can use the transposed vectors **b** and **c** to construct two indicator matrices **B** and **C** (with each element equal to 0 or 1) of order n x m that define the relationship between the technologies and components:

b_{ki} = 1 if benefit component i is delivered by technology k and else $b_{ki} = 0$

c_{kj} = 1 if cost component j is delivered by technology k and else $c_{kj} = 0$.

The total per period profits of a technology k can now be written as:

$$\pi_k = 1.1 \sum_{i=1}^{m} b_{ki} - \sum_{j=1}^{m} c_{kj}. \tag{8.1}$$

In the example we have $i, j, k = 2$ and the following matrices: $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, $\mathbf{C} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. These matrices are directly copied from table 8.1. The rows correspond to technologies while the columns represent benefit and cost components.

8.1.2 Incremental profit and the role of the installed base

The above definition of profit applies if either technology is adopted from scratch. In general, however, both the incremental benefits and costs of adopting a technology will be less than the stand-alone benefits and costs, because some benefit and cost components are already provided for by existing technologies. For example if ATMs have been adopted previously, the savings on human tellers are already realized.

The incremental costs and benefits depend on the installed base of previously adopted technologies. Let $I = \{k_1, k_2..\}$ denote this installed base of previously adopted technologies. The incremental profits are obtained by taking expression (8.1) and subtracting the benefits (adding the costs) that are already covered by at least one technology in the installed base I. The two applicability matrices **B** and **C** change: certain costs and benefits are no longer applicable because they have already been provided for. Let $\mathbf{B}^I$ and $\mathbf{C}^I$ denote these adapted applicability matrices, defined as

$$\begin{aligned} b^I_{ki} &= b_{ki}\left(1 - \max_{m \in I}\{b_{mi}\}\right) \\ c^I_{kj} &= c_{kj}\left(1 - \max_{m \in I}\{c_{mj}\}\right). \end{aligned}$$

Formally the incremental profits are now equal to:

$$\begin{aligned}\pi_k^I &= 1.1\sum_{i=1}^{m} b_{ki}^I - \sum_{j=1}^{m} c_{kj}^I \\ &= 1.1\sum_{i=1}^{m} b_{ki}\left(1 - \max_{m\in I}\{b_{mi}\}\right) - \sum_{j=1}^{m} c_{kj}\left(1 - \max_{m\in I}\{c_{mj}\}\right). \quad (8.2)\end{aligned}$$

If we want to calculate $\pi_{POS}^{I=\{ATM\}}$, i.e. the incremental profit of adopting POS with an installed base consisting of ATM, the variables would work out as follows:

$$\begin{aligned}I &= \{ATM\} \\ \mathbf{C}^{I=\{ATM\}} &= \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \mathbf{B}^{I=\{ATM\}} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.\end{aligned}$$

The first row of both matrices, which corresponds to ATM is 0 because ATM technology is in the installed base, and thus adopting it yields neither incremental costs nor benefits. The second row of the two matrices corresponds to POS, and since the two technologies share no cost elements, the bottom row of $\mathbf{C}^I$ is identical to the bottom row of $\mathbf{C}$ in (8.1). The bottom row of $\mathbf{B}^I$ is different however, since both technologies provide teller savings. Hence b_{21}^I (the bottom left element of $\mathbf{B}^I$) is now 0: teller savings are already provided by ATM. We can now calculate the incremental profit of ATM on top of POS:

$$\pi_{POS}^{I=\{ATM\}} = 1.1\sum_{i=1}^{2} b_{POS,i}^{I=\{ATM\}} - \sum_{j=1}^{2} c_{POS,j}^{I=\{ATM\}} = 1.1 - 1 = 0.1.$$

Using a similar approach the incremental profit of POS on top of ATM is:

$$\pi_{ATM}^{I=\{POS\}} = -1.$$

The total profit of adopting both technologies is then equal to:

$$\pi_{\{ATM,POS\}} = \pi_{ATM} + \pi_{POS}^{I=\{ATM\}} = \pi_{POS} + \pi_{ATM}^{I=\{POS\}} = 0.2.$$

Here $\pi_{\{ATM,POS\}}$ denotes profits after adoption of ATM and POS. A few interesting observations can be made, even from this very simple example.

1. *Arrival order matters.* Suppose POS arrives first; if ATM arrives later, it will not be adopted: adopting ATM on top of POS offers no extra benefits while there are extra costs. On the other hand, if ATM arrives first, POS may be adopted as well.

2. *Past adoption choices affect future adoption decisions.* Suppose one country adopts POS while the other doesn't, for whatever reason. If ATM arrives later, the second country adopts it, while the first does not.

3. *The economic impact can be substantial.* If a country directly adopts POS it will end up with a profit of 1.2, much higher than the 0.2 it gets if it adopts ATM and then POS.

The examples assume that the ATM technology cannot be simply dropped once POS comes along. The most obvious explanation would be sunk ATM costs. However, I find it hard to assume these (or any) costs are truly sunk over the long or medium term. I would rather offer two alternative explanations. In the first place, substituting one technology by another requires even more coordination than just adopting a new technology. And as earlier chapters of this thesis have argued, even coordination to just adopt a single new technology is challenge. Secondly, as the number of technologies rises, it may no longer be a simple matter of dropping one technology and adopting another. Rather, it may imply dropping several technologies, while simultaneously adopting several new ones. This makes coordination an order of magnitude more difficult, especially if interests are not perfectly aligned. I will return to this issue when I discuss the results of the model in the last section of this chapter.

8.1.3 Definition of equilibrium and regret

Before we can use the above model to answer the three questions posed at the beginning of this chapter, we need to define what we mean by multiple equilibria, and how we measure the economic impact of any resulting path dependence. I therefore introduce the following definitions.

Point: *A set of costs and benefits components that can be reached by adopting a set of technologies.*

The example has four points, corresponding to $I = \emptyset$, $\{POS\}$, $\{ATM\}$ and $\{ATM, POS\}$.

Equilibrium point: *A point that yields a profit, while this profit cannot be increased further by adopting an additional technology.* Let that point be reached by adopting a set of technologies $I = \{k_1, k_2..\}$, then the following must hold: $\pi_{\{k_1,k_2..\}} > 0$ and $\pi_k^I \leq 0$ for all k: profit after adopting $k_1, k_2..$ must be positive, and it is not possible to further improve profits by adopting another technology.

The example has two such equilibrium points, corresponding to $I = \{POS\}$ and $I = \{ATM, POS\}$ respectively.[3]

Path: *An ordered set of technologies, such that each technology yields an incremental profit given an installed base consisting of all previous technologies*: $P = \{k_{a1}, k_{a2}, ..., k_{av}\}$ such that $\pi^I_{ai} > 0$, where $I = \{k_{a1}, ..k_{a(i-1)}\}$ for all $i \leq v$ (and $I = \emptyset$ *if* $v = 1$).

The example has three such paths: $\{ATM\}$, $\{POS\}$, and $\{ATM, POS\}$.[4]

Accessible equilibrium point: *An equilibrium point that has at least one path leading to it.* Both equilibrium points of the example are accessible

Regret: *The difference in profit between an accessible equilibrium point and the maximum accessible profit that could have been reached through a different adoption sequence.*

Potential regret: *The maximum possible regret, i.e. the difference in profit between the accessible equilibrium points with the lowest and highest profit.*

In the example the potential regret is 1, corresponding to the difference between $\pi_{\{ATM,POS\}} = 0.2$ and $\pi_{\{POS\}} = 1.2$.

Think of the technology landscape as a mountain range. Each combination of technologies defines a *point*, i.e. an installed base consisting of a number of benefit and cost components. The total profit of such a point (installed benefit components minus installed cost components) can be thought of as the *height* of that point. The equilibrium points can be thought of as *peaks*: there are no nearby 'higher points' that can be reached by adopting one other technology. As in the mountains, this landscape can be accessed through *paths*. While each path leads to a point (by definition) not all paths will lead to peaks (equilibrium points). For example the path $P = \{ATM\}$ does not lead to an equilibrium point. Conversely, some equilibrium points have no path leading to them, for example because all the potential paths leading up to them include the adoption of one or more technologies with a negative incremental profit.[5]

---

[3] $I = \{ATM\}$ is not an equilibrium point since POS can be profitably adopted on top of that installed base, while $I = \emptyset$ is not an equilibrium point since profits are equal to zero.

[4] $P = \{POS, ATM\}$ is not a path since $\pi^{I=\{POS\}}_{ATM} < 0$.

[5] For an interesting example of an inaccessible equilibrium point, consider the following case:

$$\mathbf{C} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{8.3}$$

Both technologies have a profit of $\pi_k = -0.9$, but their joint adoption yields $\pi^{I=\{2\}}_1 = \pi^{I=\{1\}}_2 = 0.2$. Thus there are no paths, and one non-zero equilibrium point. It turns out that cases where the profit of an inaccessible equilibrium point is higher than the profit of the highest reachable point are rare. And if they occur, the economic significance is limited.

The introduction to this chapter posed three questions regarding the effect of the installed base on technology succession. I can now rephrase these questions in the context of the model:

1. Does the model have multiple accessible equilibrium points? If so, how often does this occur?

2. What is the maximum possible regret, i.e. the difference between the accessible equilibrium points with the highest and lowest profits?

3. What properties of the initial structure of the model affect the occurrence of multiple accessible points and the maximum regret? With the initial structure of the model I mean the parameters m and n and the specific structure of the two matrices **B** and **C**.

The next proposition answers the first part of the first question affirmatively, using the example of ATM and POS.

Proposition 8.1 *For all* m,n ≥ 2 *there are benefit and cost matrices* **B** *and* **C** *such that path dependence with positive potential regret can occur.*

Proof. The ATM/POS example gives an example of multiple equilibria and positive potential regret for the case where $m = n = 2$. For larger m and n similar examples can be constructed. The simplest way is of course to expand the same matrices from the 2 by 2 ATM/POS example to larger m and n by filling the other positions in the matrix with 0's. A more interesting example for $m = n$ is:

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 & .. & 0 \\ 1 & 1 & 0 & .. & 0 \\ 1 & 1 & 1 & .. & 0 \\ .. & .. & .. & .. & .. \\ 1 & 1 & 1 & .. & 1 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & .. & 0 \\ 0 & 1 & 0 & .. & 0 \\ 0 & 0 & 1 & .. & 0 \\ .. & .. & .. & .. & .. \\ 0 & 0 & 0 & .. & 1 \end{pmatrix}.$$

If these technologies are adopted starting from the first row down to the last, all of them will be adopted, and profit will be equal to $1.1n - n = 0.1n$. However if the last technology is adopted first, no other technology will be adopted (all benefit components are already provided) and profit will be equal to $1.1n - 1$. Thus the potential regret is equal to $n - 1$. If we express the potential regret as a percentage of the maximum profit, it becomes:

$$\frac{n-1}{1.1n-1} = \frac{1.1n-1}{1.1n-1} - \frac{0.1n}{1.1n-1}$$

This expression goes to $1 - \frac{0.1}{1.1} \approx 91\%$ as $n \to \infty$. A more realistic example (again for $m = n$) could be:

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 & .. & 0 & 0 & 0 \\ 0 & 1 & 0 & .. & 0 & 0 & 0 \\ 0 & 0 & 1 & .. & 0 & 0 & 0 \\ .. & .. & .. & .. & .. & .. & .. \\ 0 & 0 & 0 & .. & 0 & 1 & 0 \\ 1 & 0 & 0 & .. & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & .. & 0 & 0 & 0 \\ 1 & 1 & 0 & .. & 0 & 0 & 0 \\ 0 & 1 & 1 & .. & 0 & 0 & 0 \\ .. & .. & .. & .. & .. & .. & .. \\ 0 & 0 & 0 & .. & 1 & 1 & 0 \\ 0 & 0 & 0 & .. & 0 & 0 & 1 \end{pmatrix}.$$

If these technologies are adopted starting from the first row down to the last, all of them except the last one will be adopted, and profit will be equal to $1.1(n-1) - (n-1) = 0.1(n-1)$. If the last technology is adopted first, no other technology is adopted (all benefit components are already provided) and profit will be equal to 0.1. Thus the potential regret is equal to $0.1(n-2)$. Expressed as a percentage of maximum profit this is:

$$\frac{0.1(n-2)}{0.1(n-1)} = 1 - 0.1\left(\frac{1}{n-1}\right)$$

and this expression goes to 1 as $n \to \infty$. ■

The next section explores the prevalence and economic significance of these multiple equilibria.

8.2 Analysis of number of equilibrium points and size of regret

The occurrence of (accessible) equilibrium points depends on the precise structure of the two matrices **B** and **C**. Given m and n, each of these matrices can have 2^{mn} different structures. There are therefore 2^{2mn} forms for **B** and **C** combined. Below I first analyze all 256 situations of the case $m = n = 2$, before extending these results to larger values of m and n.

8.2.1 Analysis of 2 by 2 case

Each of the 256 situations corresponds to a specific combination of 2 cost and 2 benefit elements for 2 technologies. All of these 256 situations have either 0, 1 or 2 accessible equilibrium points. The breakdown is as follows.

- In 36 cases neither of the 2 technologies yields a profit if adopted standalone.[6] In each of these situations there is no accessible equilibrium point. These 36 cases include 2 cases where there is an inaccessible equilibrium point; these correspond to the example given in (8.3) and its mirror image.

- There are another 176 cases where there is only one (non-zero) accessible equilibrium point. These include two types of situations:

  - All 120 cases where there is only one technology that yields a profit on a standalone basis.[7]

  - 56 cases where there are two profitable technologies (standalone) such that $\pi_2^{I=\{1\}} > 0$ and $\pi_1^{I=\{2\}} > 0$, i.e. after adopting one technology it is still profitable to adopt the other.[8]

- Finally there are 44 cases with 2 accessible endpoints. These include the situation used in the ATM/POS example described earlier.

If we assume that all 256 cases are equally probable, there is therefore a $44/256 \approx 17.2\%$ probability that path dependence is a relevant phenomenon.

Can multiple equilibrium points indeed lead to substantial differences in profit? The answer is yes, potential regret occurs in the vast majority of these 44 cases. The breakdown is as follows: 6 of the 44 cases yield the same profit for both accessible equilibrium points. In the others there is an average difference of 1.04 between the two points. Since the maximum possible profit is 2.2, these differences in profit are substantial: in a majority of cases the difference is half or more of the maximum possible profit.

---

[6] To see this, note that each technology is unprofitable if the number of cost elements is larger than the number of benefit elements. Each technology has 16 possible combinations of cost and benefit elements. Of these, 6 are unprofitable: 4 cases where there are 0 benefit elements, and 2 cases where there is one benefit element and two cost elements. There are therefore $6^2 = 36$ cases where both technologies are unprofitable if adopted standalone.

[7] Each technology is profitable standalone if the number of benefit elements provided exceeds the cost elements required. This happens in 10 out of 16 cases. In the other 6 out of 16 cases a single technology is unprofitable standalone. Thus there are 10x6+6x10 cases where just one technology is profitable (one of the two has to be profitable, the other unprofitable).

[8] As an example consider

$$\mathbf{C} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

i.e. the technologies share neither cost nor benefit components.

8.2.2 Extension to larger m and n

For larger m and n the number of combinations quickly becomes unmanageable. I have therefore used Monte Carlo simulation to analyze these cases. For each combination of m and n I generated 10,000-100,000 combinations of the matrices **B** and **C**.[9] For each of these combinations I calculated the following variables:

$$\begin{aligned} a(m,n) &: \text{the number of accessible equilibrium points} \\ d(m,n) &: \text{the potential regret.} \end{aligned}$$

To determine how 'common' the occurrence of multiple equilibria is, I calculated $P[a(m,n) > 1]$, i.e. the probability that a specific combination of matrices **B** and **C** enable multiple accessible equilibria, given m and n. As we saw in the 2 by 2 case: $P[a(2,2) > 1] = 44/256 \approx 17.2\%$, assuming all cases occur with equal probability. This last assumption implies that the matrices **B** and **C** contain on average an equal number of 0's and 1's. While this is fine for smaller m and n, intuitively the matrices should become 'emptier' for larger values. To take this into account, I introduce the parameter p, denoting the probability that a single element of **B** and **C** is equal to 1. For the remainder of this section I will set $p = \frac{1}{2}$, corresponding to the situation where all possible combinations are equally probable. The simulation in the next section uses $m = n = 10$, and I will then set a somewhat lower value, $p = 0.3$.Figure 8.1 shows the variable $P[a(m,n) > 1]$ for various values of m and n, with $p = \frac{1}{2}$. A few observations stand out. First, multiple accessible equilibrium points exist with near certainty if $n > 5$. Second the effect of increasing m, the number of components, is very modest. The following conjecture derives analytical approximations for the average number of accessible equilibrium points, $\overline{a}(m,n)$, and $P[a(m,n) > 1]$.

Proposition 8.2 *For* m = 1 .. 10, *the value of the following variables can be approximated: the probability that an individual technology is profitable, the average number of accessible equilibrium points, and the probability of multiple equilibria. The approximations are as follows:*

1. *The probability that an individual technology is profitable is equal to:*

$$P(\pi_k > 0) = \frac{1}{2} + \frac{1}{2}\left[\sum_{k=1}^{m} \binom{m}{k} p^k (1-p)^{m-k}\right]^2 - \frac{1}{2}p^{2m}. \tag{8.4}$$

[9] 100,000 combinations for $n \leq 4$, 10,000 combinations for larger n.

2. *Given* m *and* n, *and* p= $\frac{1}{2}$, *the average number of accessible equilibrium points is approximately equal to:*

$$\overline{a}(n,m) \approx 2^{q+1}\sqrt{\frac{2q-3}{2\pi q}} \text{ with} \tag{8.5}$$

$$q = \frac{2}{3}nP(\pi_k > 0) \text{ where (8.4) defines } P(\pi_k > 0).$$

3. *Given* m *and* n, *the probability of multiple equilibria can be approximated by:*

$$P[a(m,n) > 1] \approx 1 - F_{Weibull(\alpha,\beta)}(1), \tag{8.6}$$

where $F_{Weibull(\alpha,\beta)}$ *is the cumulative Weibull distribution with parameters:* $\alpha = \frac{\overline{a}(n,m)}{\Gamma(1+\frac{1}{\beta})}$ *and* $\beta = 2.5$.[10]

Proof. See appendix. ■

Figure 8.2 shows $\overline{d}(m,n)|[a(m,n) > 1]$, i.e. the average potential regret given that there are multiple accessible equilibria. As before, potential regret is defined as the difference in profit between the accessible equilibrium points with the highest and the lowest profit respectively. For the purpose of the graph I have expressed average potential regret in relative terms, as a percentage of the maximum accessible profit. This relative potential regret increases with both m and n.

In summary, I find that multiple equilibria exist with near certainty, even for relatively simple cases, where there are 5 to 10 technologies and less than 10 benefit and cost components. The economic impact of this phenomenon is significant, because the average potential regret is 50% or more of the profit (or welfare) of the optimal equilibrium. This means that even in relatively simple technology environments, adoption order matters; historic coincidences may well have a significant impact on the eventual profit and welfare levels. To put it in jargon: technology adoption is a non-ergodic process.

8.2.3 Effect of cost and benefit structure on occurrence of multiple equilibria

So far I have focused on the effect of the number of technologies (n) and the number of components (m) on the occurrence of multiple equilibria and the

[10] Here Γ denotes the Gamma function: $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dt$. α is the scale parameter of the Weibull distribution and β is the shape parameter. Since the mean of a Weibull (α,β) function is equal to $\alpha\Gamma(1+\frac{1}{\beta})$, we get the best Weibull fit by taking $\alpha = \frac{mean}{\Gamma(1+\frac{1}{\beta})}$.

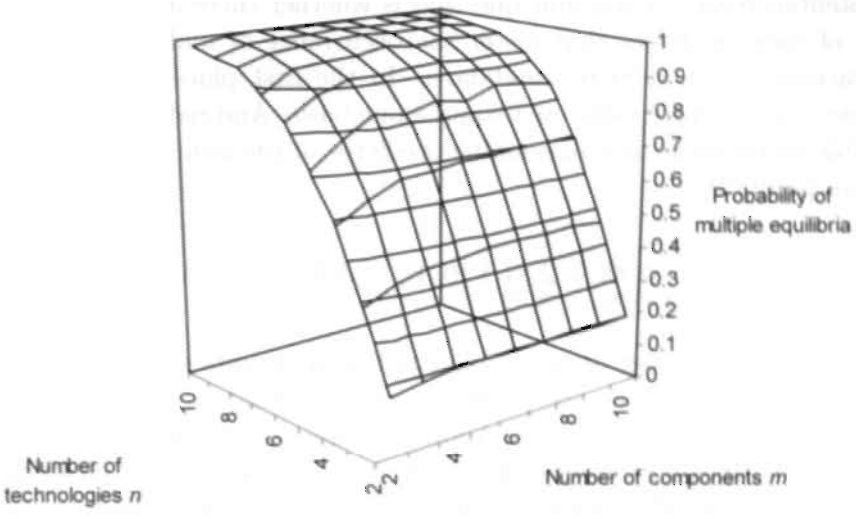

Figure 8.1 Probability of multiple equilibria given number of technologies n, and number of cost and benefit components m
Note: the graph represents averages for 100,000 runs of the model (10,000 for $n \geq 5$).

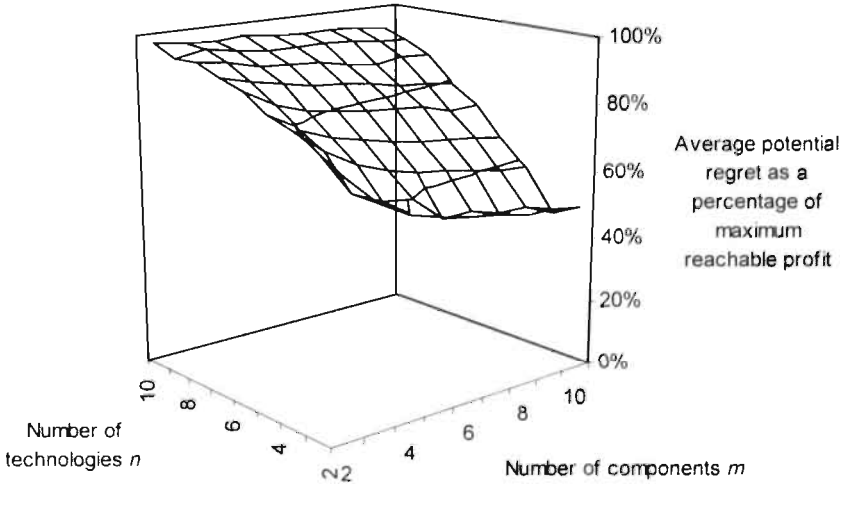

Figure 8.2 Average potential regret (difference between equilibrium points with highest and lowest profit) if multiple equilibria exist, given number of technologies m and number of technologies n
Note: the graph represents averages for 10,000 runs of the model (100,000 for $n \leq 4$).

size of potential regret. A relevant question is whether there are things in the structure of these matrices that foster the occurrence of multiple equilibria. Several variables seem obvious candidates. In the first place the number of technologies that are profitable on a stand alone basis. And indeed the number of profitable technologies is a reasonable predictor of the number of accessible equilibrium points:[11]

$$\underset{(t=5.29)}{\text{EQUIL} = -2.88} + \underset{(31.8)}{2.84\ \text{PROF}}, \qquad \text{Adj. } R^2 = 50.2\%.$$

Here EQUIL is the number of accessible equilibrium points and PROF is the number of technologies that are profitable on a standalone basis. However, beyond this, things become difficult. Adding variables that describe the structure of **B** and **C** does not increase the fit of the regression in any meaningful way.[12] Indeed, very small changes in either matrix can cause a large jump in the number of accessible equilibrium points. To illustrate this consider figure 8.3.It shows the number of accessible equilibrium points for 100 iterations of a 10 by 10 version of the model, with $p = 0.5$. Throughout the iterations **B** was kept the same, while **C** was changed in a minimal way: in each iteration I swapped two randomly selected cost elements that belong to the same randomly selected technology.[13] This procedure keeps the number of standalone profitable technologies constant. In spite of that, the number of accessible equilibrium points changes constantly. In fact there are only 13 iterations where it stays the same. This suggests that relationship between the structure of the matrices and outcomes is complex.

8.3 Impact of initial differences in installed base

The previous section showed that path dependence and significant regret are almost always possible in a landscape with 5 or more technologies: depending on the order of adoption of the technologies, different equilibrium points with different profits may be reached.

---

[11] Regression is based on 1000 Monte Carlo simulations for $m = n = 10$ and $p = 0.5$.

[12] Variables tested include: (1) the standard deviation of the sum totals of the matrices measuring whether the elements are evenly spread or concentrated in a few columns, thereby creating overlap in costs or benefits; (2) the total number of elements in the matrices; and (3) the number of columns with a particularly low or high number of elements.

[13] The only requirement was that the two elements were different (otherwise the swap does not change anything). For example in iteration 4, the cost vector of technology 4 changed from {**1**000100011} to {**01**00100011}; the affected elements are in boldface.

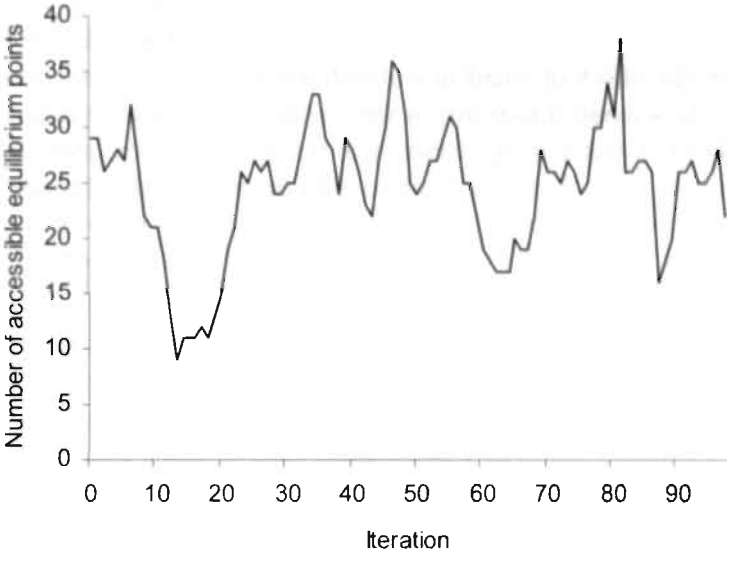

Figure 8.3 Number of accessible equilibrium points for 100 iterations where each subsequent iteration changes one cost component of matrix C, with B and number of profitable technologies constant ($m = n = 10, p = 0.5$)

The question in the context of this thesis however, is whether countries indeed reach different equilibrium points. This is not obvious. Technologies tend to become available in a certain order. It seems reasonable to assume countries will adopt these technologies in the same order that they arrive in, as long as their adoption is incrementally profitable. This section examines what happens if countries start with (small) differences in the initial installed base of technologies but are otherwise confronted with technologies whose arrival order is the same for all countries; do these countries select different technologies and follow different technology paths? And if so, does this lead to significant differences in profits?

I first expand the model of the previous section by adding an arrival order of technologies and a decision rule for countries to adopt arriving technologies. I then explore the effect of small initial differences in the installed base. Overall I find that: (1) a small difference in the initial installed base leads to different paths in about a third of all cases; (2) these differences lead to an average profit difference of 25%; (3) it pays to wait: countries that delay adoption of technologies tend to achieve a higher eventual profit; (4) globalization, which makes adopting technologies of other countries more attractive, has only a modest positive influence on technology convergence across countries; and (5) these results are robust against changes in the main parameters of the model.

8.3.1 Mechanics of the simulation model and sample run

To simulate 'technology paths' I use the model described in the previous section, with the following parameters.

1. *Technologies and components.* I take 10 components for both costs and benefits ($m = 10$). Each of the benefit components has a value of 1.1 and each of the cost components has a value of 1.[14] I then make 10 technologies ($n = 10$), each composed of a random selection of these cost and benefit components; I do this by randomly generating the matrices $\mathbf{B}$ and $\mathbf{C}$ of the model in the previous section with $p = 0.3$, i.e. each element of $\mathbf{B}$ and $\mathbf{C}$ has a 0.3 probability of being 1, and a 0.7 probability of being 0. The only restriction placed is that each technology has as least one cost component.[15]

[14] Alternatively, one can randomize the value of each cost and benefit component around 1. I have performed the analysis in this section with such a randomization in the interval [0.75-1.25]. This did not lead to materially different results.

[15] The justification for this restriction is that technologies that do not require any cost components are adopted straight away, i.e. before the first period, and are thus part of the pre-existing installed base; see point 5.

2. *Arrival order.* At the beginning of each period one cost component becomes available (the technological component is 'invented'): cost component 1 becomes available at $t = 1$, and so on, until component 10 becomes available at $t = 10$. A technology is available to firms once all its required cost components have become available: if technology k requires cost components 3 and 8, it can be adopted in period 8 or later.

3. *Adoption decision.* At the beginning of each period t, firms evaluate the incremental benefits of all available technologies, using the formulas in (8.2). The technology with the maximum positive incremental profit is adopted.

4. *Drop decision.* After each period, firms evaluate all previously adopted technologies. An adopted technology k is dropped if the incremental profits of doing so are non-negative; this is the case if many cost components can be dropped without losing too many benefits (because they continue to be provided by the remaining technologies).[16]

5. *Number of countries.* I start with two countries, but also examine the situation with 10 countries in section 8.3.2.

6. *Initial situation and country differences.* I assume a country starts with an initial installed base of just one benefit component being fulfilled, and no cost components in place. To analyze the effect of initial differences, I assume country 1 has benefit component 1 in place, country 2 has benefit component 2 in place, etc.

Before describing the results of this simulation model, I first derive some important characteristics by applying equations (8.4), (8.5), and (8.6) of proposition 8.2 for the selected parameter values ($m = n = 10, p = 0.3$):[17] Applying expression (8.4) gives the probability that an individual technology is profitable on a stand alone basis:

$$P(\pi_k > 0) = \frac{1}{2} + \frac{1}{2}\left[\sum_{k=1}^{m} \binom{m}{k} p^k (1-p)^{n-k}\right]^2 - \frac{1}{2}p^{2m} = 60.0\%.$$

[16] To avoid unnecessary technology clutter, I assume countries drop a technology even if the incremental profit of doing so is zero. This assures that the installed base is as clean as possible.

[17] The formulas change slightly because I now use $p = 0.3$ instead of $p = 0.5$ (where p is the probability that a single element of $\mathbf{B}$ and $\mathbf{C}$ is 1). In the last formula the β parameter of the Weibull distribution changes from 2.5 to 1.8, but the Weibull distribution still fits the actual (Monte Carlo based) distribution of the number of accessible equilibrium points almost perfectly.

The number of equilibrium points is obtained through expression (8.5):

$$q = \frac{2}{3}nP(\pi_k > 0) = 4, \text{ so we get}$$

$$\overline{a}(n,m) \approx 2^{q+1}\sqrt{\frac{2q-3}{2\pi q}} = 14.00 \quad .$$

Finally expression (8.6) gives the probability that there are multiple equilibria:

$$P[a(m,n) > 1] \approx 1 - F_{Weibull(\alpha,\beta)}(1) = 99.0\%,$$

$$\text{with } \alpha = \frac{14}{\Gamma(1+\frac{1}{\beta})} \text{ and } \beta = 1.8 \; .$$

Thus for each run of the model we expect there to be on average 6 profitable technologies, 14 accessible equilibrium points, while the probability that there exist multiple equilibria is 99%.

Let me describe a sample run of the model. The randomly generated technologies are shown in figure 8.4: each dot represents one technology, with its total cost per customer on the horizontal axis ($\sum_{j=1}^{m} c_{kj}$) and its profit under adoption on the vertical axis: π_k as defined in equation (8.1). In this sample run there are 5 technologies that are profitable on a stand alone basis. The 6 technologies that are actually adopted by one or both countries are shown by squares, with the reference number of the technology next to it, the other 4 technologies are shown as diamonds. Note that two technologies are adopted even though they yield a loss on a standalone basis: technologies 5 and 9; conversely another technology that is profitable standalone does not get adopted (the diamond just to the right of technology 2 in the graph).

These technologies become available over time. As the system adopts some and drops others, total profits rise. Figure 8.5 shows the development of profit per customer for two 'countries' which face the same technologies and arrival order. The countries differ only in that country 1 initially has benefit component 1 in place (being provided by previous technologies), where country 2 has benefit component 2 in place; hence both countries start with a profit of 1.1.

Figure 8.5 shows that the two countries may follow quite different adoption sequences, even though the technologies arrive in the same order (the small number next to each plot point refers to the technology that was adopted to reach that plot point, negative numbers indicate technologies that were dropped). As a result, the two countries end up with a different installed base and a significant difference in profits. Figure 8.6 gives an overview of the development of the underlying costs and benefit components; each row contains 10 positions corresponding to the cost and benefit components, and the dots

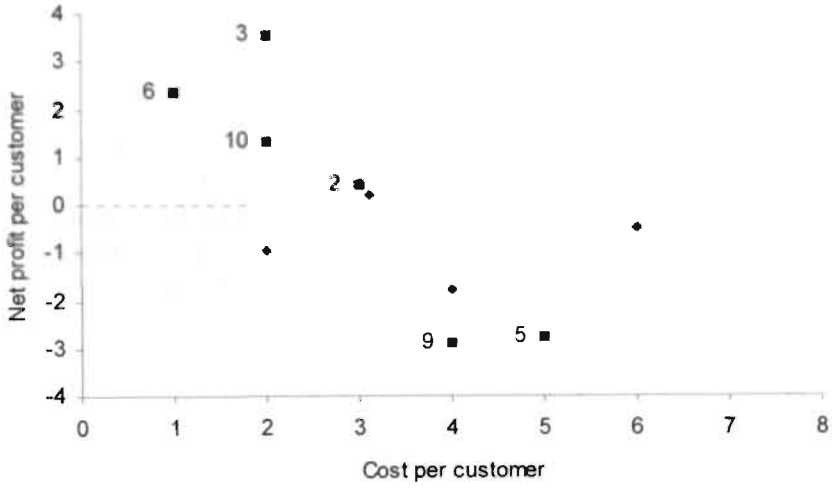

Figure 8.4 Overview of profits and costs of the technologies involved in the sample run of the simulation model

Note: each point represents a technology; squares are adopted technologies, with the serial number of the technology next to it; diamonds represent technologies that were not adopted in the sample run.

TABLE 8.2 Results for 10,000 runs of the model with 2 countries that start with a slightly different installed base

	Average outcome for both countries	Difference between countries	
		Maximum	Average
Benefit	9.66	9.9	1.03
Cost	4.46	9	1.55
Profit	5.20	4.2	0.81
Number of techn. adopted	3.43	5	1.03
Number of techn. dropped	0.50	4	0.46
Adopt different technol.	35.2%	n/a	n/a

Note: $m = n = 10, p = 0.3$. Figures in last column represent averages across cases where countries follow different paths.

represent those components that form part of the installed base for a certain period, while an empty position means the components were not part of the installed base. Note that at the start, both countries each have one (different) benefit component in place. Note also that after the final adoption, the two countries have very different cost components. They also have different benefit components, although the differences are actually quite small.

8.3.2 Convergence and divergence across countries

The above results represent only one run of the model. These results are highly dependent on the specifics of the technologies (i.e. the randomly generated matrices **B** and **C**), which in turn determine their arrival order and adoption. I have therefore calculated averages over 10,000 runs. Each run is based on the same overall parameter settings as the sample run described earlier: $m = n = 10$, $\pi_k = 1.1 \sum_i b_{ki} - \sum_j c_{kj}$ and the probability that a technology uses a specific cost or benefit element is $p = 0.3$. Country 1 always starts with benefit component 1 in place, and country 2 starts with benefit component 2 in place, etc. Other than that all countries have a clean sheet at the start. For each run **B** and **C** were randomly generated. I first discuss the results for 2 countries before expanding it to 10 countries.

Results for 2 countries

Table 8.2 summarizes the results for the two-country case. Countries adopt different technologies in a third of all cases.[18] The economic impact of this can

[18] The divergence figure in the table, 35.2%, was obtained by counting all cases where the installed bases of the two countries differ by at least one cost component; because of the

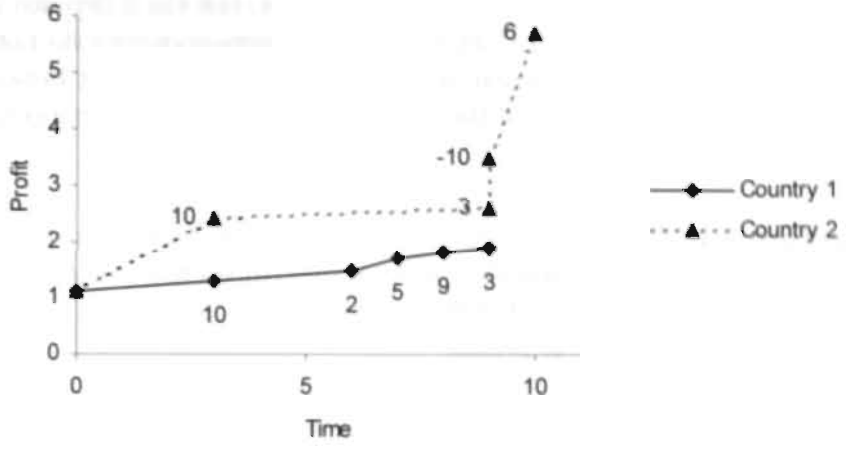

Figure 8.5 Example of different technology paths for 2 countries that start with a small difference in their installed base

Note: each point represents the adoption or dropping of a technology. The numbers next to the points are the serial numbers of the technology; negative numbers correspond to technologies that were dropped.

	Period	Cost components	Benefit components
		12345678910	12345678910
Country 1	0		.
	3	..	. . .
	6	... .	.. . . .
	7	... ...	....
	8	... .	....
	9	... .	....
Country 2	0		.
	3	..	.. . .
	9	.. ..	.. . ...
	9	..	. . ...
	10	..	.. . .

Figure 8.6 Development of installed base of cost and benefit components for 2 countries following different technology paths

Note: Dots represent cost and benefit elements in the installed base, blank spaces represent elements that are not part of the installed base.

be significant. Due to the symmetry of the initial situation, both countries end up with the same average benefits, costs and profit, but the difference in, e.g., benefits can be as high as 9.9.[19] In those cases where countries do not adopt the same technologies, the average regret, i.e. difference in profit between the two countries, is 0.81 or 15.6% of the average profit.[20] The following conjecture summarizes these findings.

Conjecture 8.1 *There is a substantial probability (35% of cases) that an initial difference in the installed base of technologies causes two countries to adopt different technologies and reach different equilibrium points. On average this causes regret equal to 15% of profits.*

Results for 10 countries

The initial differences can easily be expanded to more than two countries. I have performed 10,000 runs of the model with 10 countries, where each of these has initially one (different) benefit element in place. Table 8.3 summarizes the results. In only 21.6% of cases do all 10 countries end up with the same technology landscape.[21] In most of the other cases there are 2 or 3 groups of countries that adopt the same technologies and end up with a similar installed base of cost and benefit components (the average outcome is 2.76 different endpoints). The distribution of the number of different endpoints is best described by a Weibull distribution with $\alpha = \frac{2.76}{\Gamma(1+\frac{1}{\beta})}$ and $\beta = 1.8$. This is perhaps not surprising, since the distribution of *all* accessible equilibrium points follows a Weibull distribution with $\alpha = \frac{14}{\Gamma(1+\frac{1}{\beta})}$ and $\beta = 1.8$, where 14 was the average number of accessible equilibrium points.

The following conjecture follows from table 8.3.

---

initial differences in benefit components, countries can have small differences in the installed benefit components even if they adopted the same technologies.

[19] This maximum difference is reached in a particularly dramatic case, where the first country fails to adopt *any* technology, while the second country adopts *five*. The first available technology is not adopted by country 1 because of an overlap with benefit element 1. Because country 1 does not gain the installed cost base of that technology, it misses all subsequent adoptions.

[20] Due to difference in the initial installed base, there are differences in profit even if the countries adopt the same technologies. When two countries adopt exactly the same technologies, the difference in profit is on average 0.27 or 5.2% of the average profit.

[21] This compares to 65% for 2 countries. Note that the process does not behave like independent draws. In that case the probability of all countries ending up with the same infrastructure would be $(65\%)^9 \approx 2.1\%$. This is in line with the earlier finding that certain characteristics of **B** and **C** foster convergence: once two countries adopt the same technologies, there is an increased chance that other countries also follow the same path.

TABLE 8.3 Number of different end points reached by 10 countries starting with slightly different installed bases

Number of distinct endpoints	Occurence
1	21.6%
2	28.2
3	23.1
4	14.3
5	7.7
6	3.4
7	1.1
8	0.4
9	0.1
10	0.0

Note: average outcomes for 10,000 runs of the model with $m = n = 10, p = 0.3$.

Conjecture 8.2 *As the number of countries with different starting positions grows, the probability that at least two of them follow different paths increases: for 10 countries there is an almost 80% probability that at least two countries follow a different technology path.*

8.3.3 Sensitivity to main parameters

Sensitivity to number of components, m, and technologies, n

Figure 8.7 shows the percentage of cases where the initial difference among two countries leads to different technology paths for varying number of components (m) and number of technologies (n). This percentage lies between 20 and 40% for most values of m and n. For large m and n the percentage stabilizes around 35%. This is an interesting phenomenon: adding complexity to the system does not substantially affect the outcome.Figure 8.8 shows the average regret for the 'losing' country (the country that ends up with the lower profit). The regret is expressed as a percentage of the winning country profit, given m and n. The percentage regret is 10-20% for all m and n, while it appears to go to 10% for larger m and n.

Sensitivity to p : probability a cost or benefit component is relevant for a technology

The other important parameter is the probability that any given cost or benefit component is used/produced by a technology. In the above results this

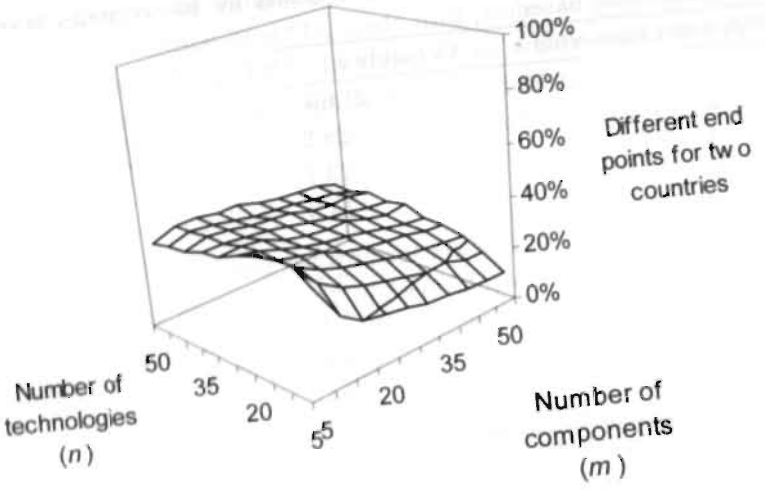

Figure 8.7 Percentage of cases where two countries end up with same technology base as a function of number of technologies (n) and components (m)
Note: averages across 10,000 runs of the model, $p = 0.3$.

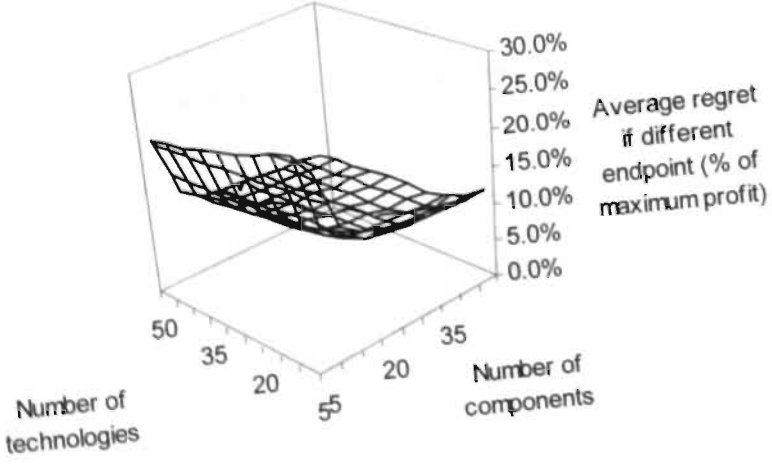

Figure 8.8 Average regret of the country that ends up in a point with lower profit a function of m and n
Note: averages across 10,000 runs of the model, $p = 0.3$.

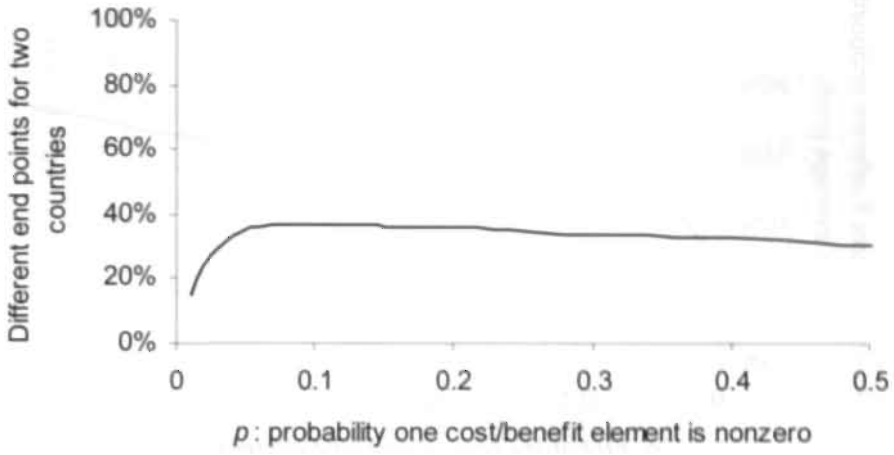

Figure 8.9 Percentage of cases where two countries end up with same technology base as a function of p: the probability single cost/benefit element is nonzero
Note: averages across 10,000 runs of the model, $m = n = 10$.

parameter was put at $p = 0.3$. Figures 8.9 and 8.10 show the sensitivity the sensitivity of the model outcome to p.

Two countries reach different end points in 30 to 40% of cases for most values of p, while average regret is 15-20%. If the probability is very low (i.e. **B** and **C** are relatively empty) there is more convergence because the chance is small that the installed base (the pre-installed benefit component of country 1 and 2) interferes with the choice of new technologies. The following conjecture summarizes these results.

Conjecture 8.3 *The model results are robust against changes in* n *(the number of technologies),* m *(the number of components) and* p *(the density of matrices* **B** *and* **C**); *an initial difference in installed base leads to different end points in 20-40% of all cases for* m *or* n > 5 *(with* p $= 0.3$*) and* p $= 0.1 - 0.5$ *(with* m $=$ n $= 10$*); in case of different outcomes average regret is 10-20%.*

8.3.4 Impact of economic factors on country convergence

Global externalities

What happens if adoption by one country makes a technology more attractive to other countries? Examples of factors inducing such an effect could be global skill spillovers, cost economies, and the fact that the technology can be more

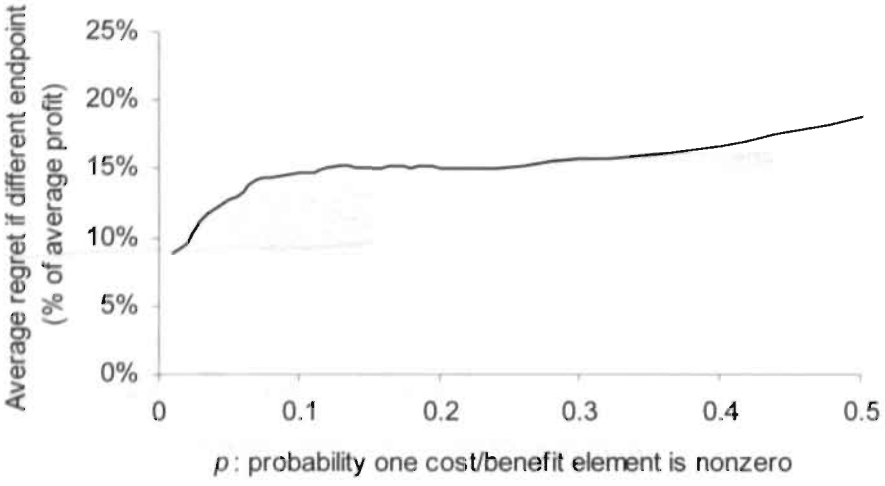

Figure 8.10 Average regret of country that ends up in a point with lower profit as a function of p : the probability that a cost/benefit component is nonzero
Note: averages across 10,000 runs of the model, $p = 0.3$.

readily used for cross-border transactions. Intuitively one would expect these global externalities to promote convergence and dampen the effect of the local legacy. I simulate this effect as follows: the incremental cost of a technology is reduced by a certain percentage if that technology is already used by the other country.[22] The results are given in figure 8.11. Obviously for zero scale effects we get divergence in 35% of cases, the same figure as in table 8.2. As is to be expected, larger scale effects lead to decreased divergence.[23] However, the percentage divergence appears to plateau at 15% as the scale effects grow increasingly large. Closer analysis reveals that these are case where one country adopts a technology whose benefits overlap with the initial installed base of the other country. Since the adoption of such a technology does not provide the second country with additional benefits, that country will not adopt the

[22] The choice for cost reduction is arbitrary. Increasing incremental benefits if a technology is used elsewhere yields the same results. Here too the curve does not drop below 15% divergence. These are cases where there is full benefit overlap between the installed base and technologies that are used in the other country.

[23] The jumps in the graph are an 'integer issue'. Cost and benefit elements have an integer value (of 1 and 1.1 respectively). Cost reductions of 45% and 90% cause a disproportionate rise in convergence.

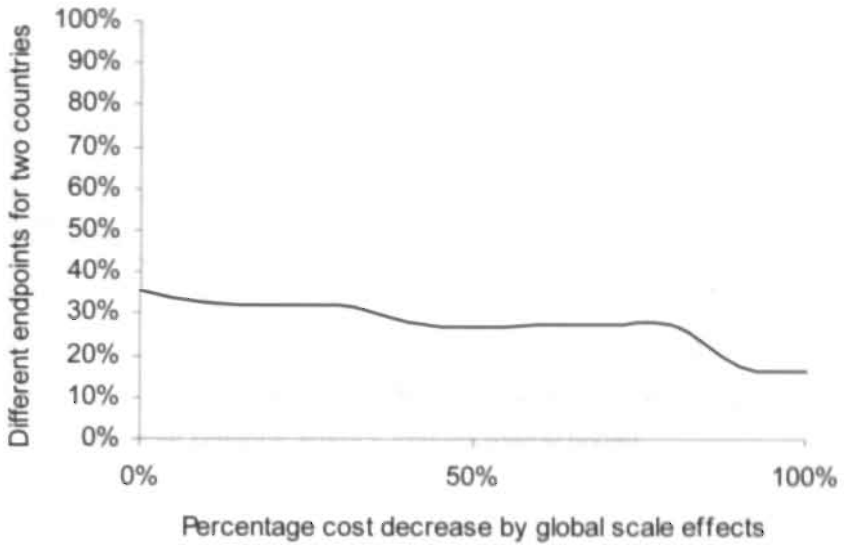

Figure 8.11 Technological convergence among two countries as a function of global scale economies Note: averages across 10,000 runs of the model, $m = n = 10, p = 0.3$

technology no matter how cheap it is. Apparently in specific cases a country legacy can be tenacious. This leads to following conjecture.

Conjecture 8.4 *The existence of global positive externalities fosters convergence. However, even for very large global scale economies, non-convergence still occurs with positive probability.*

8.3.5 Effect of late starts

What happens if a country arrives late to the party? To simulate this, I have constrained country 2 so that it cannot adopt any technology during the first $t^* - 1$ periods (and we vary t^*). The 'adoption ban' is lifted in period t^*, after which the country can choose among all technologies available at that time. To keep matters simple, I have eliminated the initial difference in benefit elements for the purpose of this exercise. Both countries start with *the same* benefit element in place in period 1 (element 1, so the results for country 1 are the same as before). Figures 8.12 and 8.13 show the results. For $t^* < 5$ there is

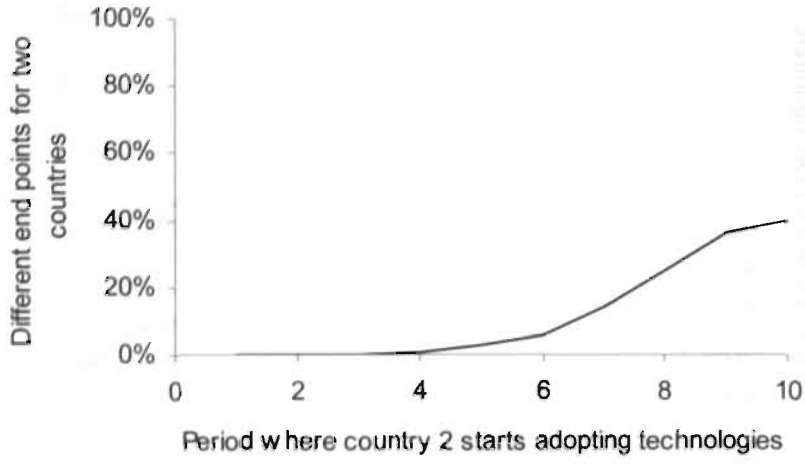

Figure 8.12 Technological divergence caused by late start of one country
Note: averages across 10,000 runs of the model, $m = n = 10, p = 0.3$. Countries start with identical base: first benefit element is 1, all else is 0.

hardly any effect.[24] For starts after period 5 ($t^* \geq 5$), divergence starts to increase, reaching about 40% for very late starts. On average, a late start leads to higher profit in case of divergence. This is however an average. In individual cases, a late start can also lead to a lower profit. Overall though, it seems that new-testamentical salvation prevails over old-testamentical damnation:

Conjecture 8.5 *On average, a late start decreases convergence, and leads to a higher profit for the delayed country. In individual cases the delayed country can end up with a lower profit.*

8.4 Incremental change versus paradigm shifts

An interesting feature of the model is the preference for incremental technologies: most technologies add only a single cost component to the cost base. Similarly, most adoptions add only one or two benefit components. As a result the average impact on profit is even smaller, with half of all adoptions

[24] This is not surprising, since even without constraints, countries only start adopting technologies in the later periods anyway. This is because a technology can only be adopted once all its cost components have become available.

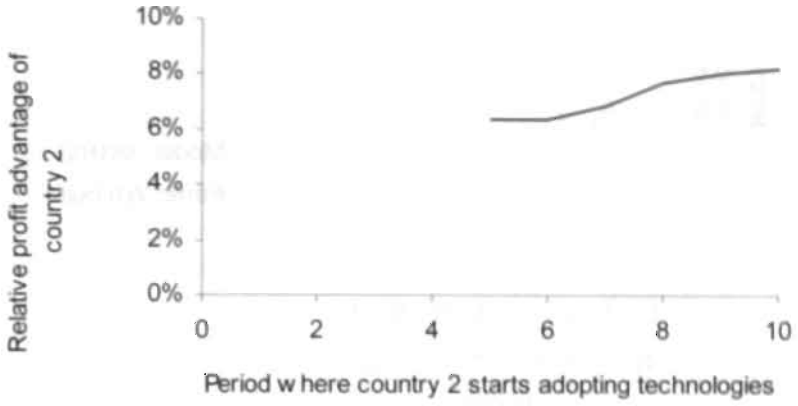

Figure 8.13 Average profit advantage in case of divergence for a country that starts late

Note: averages across 10,000 runs of the model, $m = n = 10, p = 0.3$. Countries start with identical base: first benefit element is 1, all else is 0. If country 2 starts before period 5 there are (almost) no cases where 2 countries adopt different technologies.

increasing profits by less than one.[25] Figures 8.14 and 8.15 show the frequency distribution of the increase in cost elements and in profits.For example, 53% of all adoptions increase benefits by 1 component, another 6% add 0 components. The average increase is 1.6 cost components. Similarly, most technologies increase profits by an amount in the range 0-1 (47% of cases) or 1-2 (32%). On rare occasions however, the adoption of a single technology can increase profit by more than 4 (2% of cases).

Conjecture 8.6 *There is preference for incremental technologies: technologies that add 0 or 1 cost component, 1 or 2 benefit components, and increase profits by 0-2.*

In the context of this model, both incremental innovations and paradigm shifts (as defined by Dosi, 1982) are part of the same distribution, and not driven by separate processes. The appendix contains the statistical analysis of the distribution of increases in costs and benefits. The analysis reveals something close to a Weibull (rather than a Poisson, Lognormal or Gamma) distri-

[25] This phenomenon has been called the 'sailing ship effect', and was discussed in section 3.2.2.

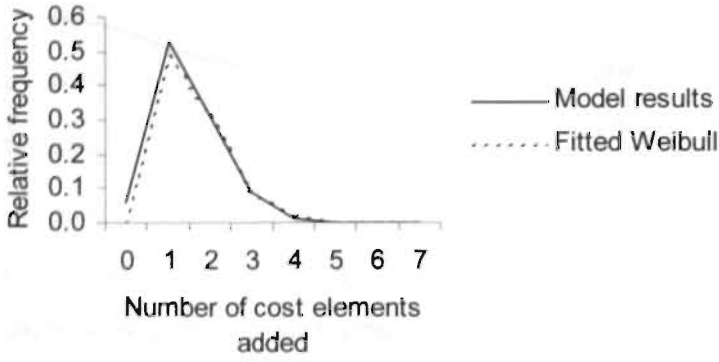

Figure 8.14 Increase in number of cost elements following each adoption of a technology
Note: averages across 10,000 runs of the model, $m = n = 10, p = 0.3$.

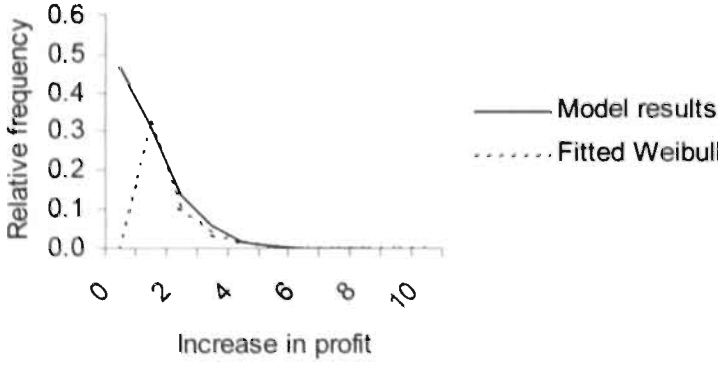

Figure 8.15 Distribution of the increase in profit caused by each adoption of a technology
Note: averages across 10,000 runs of the model, $m = n = 10, p = 0.3$.

bution, with the familiar parameters:

$$\alpha = \frac{average}{(1+\frac{1}{\beta})}, \beta = 1.8$$

Here *average* is the average increase in costs and benefits.[26] The reappearance of the Weibull is remarkable, especially with the shape parameter $\beta = 1.8$: this was the same distribution and shape parameter that described the number of accessible equilibrium points (for $p = 0.3$). Furthermore, this shape parameter is robust against changes in m. For $m = 50$, the distribution of the increase in costs and benefits is again best approximated by a Weibull distribution with $\beta = 1.8$, while the increase in profits follows a Weibull with $\beta = 1$.[27] While the specific distribution (Weibull) may be caused by the peculiarities of my model, the more general point is that the variables that measure technological progress (incremental costs, benefits or profits) follow a skewed distribution, with a bulge close to zero and a long tail of larger increases. Thus the same process causes both incremental and drastic innovation.

8.5 Discussion of results

This chapter introduced a model that deals explicitly with the impact of the installed base on the adoption of new technologies. In reviewing the model and its results, I will address two questions: (1) is the model realistic? and (2) do the outcomes have meaningful implications?

8.5.1 Is the model realistic?

The model assumes a more general concept of innovation than just a new machine or process. It could be any innovation that uses a number of 'costly things' to produce benefits. It may be a machine, a process, something organizational,

[26] The same holds for the increase in profits, except that now I get $\beta = 1$. The lack of fit for the 0-observations is an integer issue: all outcomes below 1 are counted as 0. For larger m this effect diminishes, and for $m = 50$ the fit is almost perfect, see figure D.2 in the appendix.

[27] I can only speculate about why the Weibull distribution keeps popping up. The Weibull distribution is generally used to describe the time to failure of complex systems, like engines, which does not easily translate to my model. A more promising explanation is that for $\beta = 2$ the Weibull becomes the Rayleigh distribution. It can be shown that this is the theoretical distribution of the speed of an object with two speed vectors that are independently and normally distributed with equal variance. This process turns two variables with a symmetric distribution into one variable with an asymmetric distribution (speed is positive by definition). A similar process is at work in my model where two symmetrical variables (benefits and costs are both Binomially distributed) result in one variable (incremental profit) that is positive.

a set of standards etc. While not sufficient, this is at least a requirement for a realistic model of innovation: major innovations of this century include new processes (assembly line production), new ways of organizing things (direct selling of PCs) and standards (transportation containers, the World Wide Web).[28] Furthermore, the model supposes that adoption of a technology requires embedding it into 'something that already exists', be it infrastructure, installed base etc. Again, this seems realistic. Many of the earlier examples of major innovations built on existing things. Direct selling of PCs is possible because of a delivery infrastructure (FedEx, UPS), a payment infrastructure, and arguably the Internet. The World Wide Web became viable because a large number of PCs and the Internet were already in place.

At the same time invention does play a role in the model. Invention expands the set of available (cost) components and increases the number technologies to choose from. Thus the model combines the two distinct concepts of invention of components and adoption of technologies. While inventions are available to everyone, not everyone will adopt the technologies that these inventions enable, because of differences in the installed base.

The simplicity of the model comes of course at a cost. In particular the model abstracts from firms, industry structure, and competitive behavior. Instead, it assumes that countries as a whole adopt technologies, and that cost components, once installed, are available to all other adopters of new(er) technologies. While this is clearly a strong assumption, there are a number of situations where this may apply. First, the cost components need not be physical 'things' owned by players, such as ATMs or cards in the consumer's wallet. They could also be standards or knowledge and expertise. If these are non-proprietary (as in non-proprietary standards or Marshallian spillovers), it is quite likely that they can be used for free by subsequent technologies. And even if they are proprietary, it is often in the interest of players to make them available to all.

In the model of this chapter, the players in an industry serve as a 'conduit' for technologies. Much in the same way as individual organisms are a conduit for genes, to (again) use the biology analogy. Their competition enables evolution of technologies, but does not necessarily explain its course.

8.5.2 Do the outcomes have meaningful implications?

Let me first summarize the major results of the model:

[28] Note also that many, or even most, of these innovations are not obviously patentable. This suggests that number of patents filed may not be a great proxy for the inventiveness of a country.

1. Path dependence and regret are real phenomena: multiple equilibria *can* exist for all parameters, and *will* exist with near certainty if there are more than 5 technologies. As a result, small differences in the installed base lead with positive probability (20-40%) to continued divergence between 2 countries, even as new technologies become simultaneously available to all countries. On average, such divergence leads to a 10-20% difference in welfare between 2 countries. For a larger number of countries there is of course a larger probability that at least two countries follow different technology paths. These results are robust to changes in the main parameters of the model.

2. Global economies of scale mitigate but do not eliminate this lack of convergence between countries. Even if the fact that another country has adopted a technology would *halve* the incremental cost of adoption, there is still a positive probability (15-25%) that two countries follow non-converging paths.

3. The phenomenon of the 'penalty of taking the lead' (also known as 'the incumbents disadvantage') occurs naturally: a country that fails to adopt the first available technologies, generally gains later on; the late comer has more technologies to choose from, and it can adopt any technology, while the incumbent generally first has to drop some earlier technologies.

4. Incremental and breakthrough advances in technology are part of the same underlying distribution: 'paradigm shifts' are the tail of a (Weibull) distribution that includes incremental (or even no) advances at the other end.

Outcome 1 and 2 suggest that if there is such a thing as local technology landscape, this locality does not disappear easily. Even with the arrival of globally available technologies, the local nature may persist. Even in a perfectly rational and transparent world, some countries may adopt new technologies while others may not.

Outcome 3 confirms that early adoption can create a legacy that turns into a burden. This is of course caused by some form of switching cost. In the model these switching barriers arise because I assume old technologies are only dropped if such dropping is incrementally profitable. There is no explicit trade-off between the profits of an old technology and a new technology. There are two reasons why this is realistic. First, there are often social implications to dropping a technology; these make it hard to drop a technology that is incrementally profitable, even though an even more profitable alternative is available. Second, there is seldom a clear trade-off between an old and a new

technology: to reach a higher point a country may have to backtrack its path to a suboptimal top by dropping several technologies and then take another path by adopting several new ones. A quite complicated process involving interlocking technologies, especially if the technologies are used throughout an industry.

Finally outcome 4 suggests that while 'paradigm shifts' may be a useful term to describe large technological advances, one should stop there. 'Paradigm shifts' are not inherently different from other innovations, and there is no reason why they should occur every 50 years.

Let me conclude by suggesting some areas for further research. In the first place, it would be interesting to explore the length of the paths between the various peaks: how difficult is it to get from a suboptimal equilibrium to the optimal one: how many technologies need to be changed (either dropped or adopted)? In the second place, the model has a finite time-horizon, with a clear starting point and an end point that is determined by parameter m. It would be interesting to adapt the model so that it can run forever, i.e. with a time horizon that is independent of key parameters. Thirdly, the model is very simple: what happens if complexity is increased, for example by allowing cost and benefit elements of varying value?

8.6 Chapter appendix: list of symbols used

Parameters:

b^k	:	Benefit indicator vector: $b^k_j = 1$ if benef. compon. $j$is used by techn. k
$\mathbf{B}$	:	Benefit indicator matrix: $b_{ik} = 1$ if ben.compon. $i$is delivered by techn. k
c^k	:	Cost indicator vector: $c^k_j = 1$if cost compon. $j$is used by technology k
$\mathbf{C}$	:	Cost indicator matrix; $c_{jk} = 1$if cost compon. $j$is used by technology k
i	:	Benefit components, $i = 1..m$
j	:	Cost components, $j = 1..m$
k	:	Technologies, $k = 1..n$
m	:	Number of benefit components and cost components
n	:	Number of technologies
p	:	probability that an individual element of $\mathbf{B}$or $\mathbf{C}$is equal to 1

Variables that depend on the state of the system:

$a(m,n)$: number of accessible equilibrium points, given $m$and n

$d(m,n)$: potential regret, given $m$and n

$\mathbf{B}^I$: Benefit indicator matrix given installed base I

$\mathbf{C}^I$: Cost indicator matrix given installed base I;

I : Installed base. $I = \{k_1, k_2, ..\}$if technologies $k_1, k_2, ..$have been adopted

π_k : standalone profit of technology k

π_k^I : incremental profit of technology $k$given installed base I

t^* : period after which a country may start adopting technologies

9

Case 3: Electronic money, Internet and Mobile payments

If technologies can be compared to life forms (as the evolutionary approach to technological change will have it) then the late 1990s witnessed something like a Cambian explosion of payment technologies. All of them targeted transactions over the Internet and other remote channels such as interactive TV, mobile phones etc. Some had the ambition of changing the very nature of money, envisioning global currencies no longer controlled by (central) banks.[1] By 2003 most were defunct. The vast majority of remote transactions are currently made using traditional payment instruments like credit cards, and (central) banks appear in full control of the payments business. This chapter tries to apply the models of the previous chapters to remote payments to see if it can help explain what happened.

Compared to the earlier two cases of chapter 5 and 6, the ambition of this third case is much more modest. It should be viewed as an illustration rather than as an application of the model in chapter 8. I will not try to estimate any of the model parameters, or apply any of the formulas. Instead, I will look for evidence that the basic mechanism of the previous chapter played a role in the (non-) adoption of new payment systems. It will come as no surprise that I find that this is indeed the case.

I first describe the major contending systems, and how they fared, followed by a review of what the existing literature has to say about them. Using the model in the previous chapter I then try to establish the links (shared infrastructures) between these new instruments and the existing base.

9.1 The facts: what happened

9.1.1 Description of main contenders

An extensive taxonomy is proposed by Böhle and Krueger (2001), who also classify more than 160 individual schemes in both the US and Europe. Böhle and Krueger (2001) distinguish the following categories.

[1] E.g. Chaum (1992).

1. *Enhanced access to offline payment instruments.* The products in this category try to enable the secure online (or phone) use of traditional instruments like checks, cards and transfers.[2] Without such enhancement, these offline instruments pose several risks to both consumer and merchant, if used online or over the phone. If a customer gives his card number by phone or Internet, he does not sign for the transaction. This means that he can always deny the transaction later, in which case the merchant does not get his money. If used over the Internet there is always the risk that the card number is captured by someone else and used fraudulently. If a customer uses a credit transfer for Internet or phone purchase, the merchant bears the risk that the customer does not pay. Alternatively, the merchant can require payment before the goods are shipped, but this severely delays the whole transaction. The schemes in this category generally provide a way to prevent these risks. For example I-pay in the Netherlands provides encryption at both ends of the transaction to safely get a credit transfer instruction, or a credit card instruction from the buyer to his bank with a confirmation to the merchant. I-Pay and several others rely on Secure Electronic Transaction protocol (SET) a proposed standard for remote payments that provides an encryption standard as well as methodology whereby a merchant would get payment confirmation without seeing the credit card number or bank account number of the customer.

2. *E-wallets/Virtual accounts.* These services are essentially aggregators of small transactions. For example, they will deduct $ 10 from a credit card and put it in a 'wallet' on a remote server. The wallet can then be used to make many small payments (in the range $0.25-5.00) without the need to go through the costly process of credit card authorization and booking. The card data are stored in the wallet, both for replenishing the 'small cash' stock and to be used directly for larger payments. CyberCash, with its CyberCoins, as well as PayPal fall in this category. PayPal allows anyone with a credit card and an E-mail address to set up an account. Payments can then be made to participating merchants or other private account holders. In fact, payments can be made to any E-mail address; if the recipient has no account, PayPal will automatically set one up.

[2] In the original framework of Böhle and Krueger (2001) credit cards are a separate category from access products, since technically they do not access existing money. I have however put them under access products, since most applications that enable the use of existing instruments on the Internet, include the use of credit cards.

The money can then be credited to his credit card or be used to make a payment to another account holder.

3. *Prepaid products.* These come in several varieties.

   - Single purpose schemes, like prepaid telephone cards or public transportation cards. A consumer buys a card at a given price, which enables him to purchases several units of service, such as telephone 'clicks', bus or subway rides, etc. The card is typically discarded after it is used up. Some of these schemes have tried to expand usage beyond their original purpose, enabling for example the use of public transportation prepaid cards to buy newspapers etc. In general these attempts have been blocked by the regulator: the issuers of these cards would be issuing real money and therefore they would need a bank license. The Dutch PTT circumvented this by setting up a joint venture with Postbank to transform its pre-paid technology to a real electronic purse called Chipper.[3]
   - Electronic purses like Proton (Belgium) and Mondex (UK). Also known as smart cards or chip cards, these schemes store money on a card to be used for purchases at stores, vending machines, over the phone or through the Internet (requiring a card reader).[4] The balance on the card represents real money, so there is no need to check the customer balance through his bank.[5] In addition, customer authentication, using either signature or PIN, is not required; this further reduces the throughput time and cost compared to debit and credit cards. In the period 1995-1997 a large number of such schemes were tested or rolled out. Good (1997) gives an overview of more than 20 pilots and rollouts. High profile tests were the Atlanta Olympics, the Mondex pilot in Swindon (UK) and the pilot in the Upper West Side of Manhattan. Large scale launches were

[3] Launched in 1997, it competed head-on with a scheme of the commercial banks, Chip-Knip; in 2001 Chipper was merged into Chipknip.

[4] Formally smart cards are plastic cards with a chip on it. These serve a much wider set of applications including the SIM card in a GSM phone, or credit and debit cards equipped with an EMV chip. Nevertheless in much of the payment literature the term is used to describe cards with an electronic purse.

[5] There is a technical difference between Mondex and most of the others. The bits and bytes on a Mondex card were to represent real currency, redeemable against cash. Most of the other schemes used shadow accounting, where the amount of money loaded on a card was blocked on the checking account. As a result Mondex was able to support person-to-person transfers, where most other schemes were not. For a description of Mondex and the policy issues it raises, see Stalder and Clement (1998).

made in Belgium (Proton), Netherlands (Chipper and Chipknip) and Germany (Geldkarte).

- Network money like CyberCash and DigiCash (both US). Here the bits and bytes on a computer (and/or chip as in the Mondex scheme) represent real money for the bearer. Unlike coins and notes, they can be split into any denomination. And because transaction costs consist of a few computer calculations, they would be fit for 'micro payments', where one would pay say a cent per newspaper page viewed. All of this was to be made possible through software encryption technology. Indeed DigiCash was the brainchild of a cryptologist (David Chaum).
- Dedicated prepaid accounts like gift currencies, where a third party (parents, friends) loads the account to be used at selected shops. An example is VisaBuxx, essentially a Visa card with a limited spending amount on it, to be used by children for purchases on e.g. the Internet.

4. *Money surrogates* like e-vouchers, e-bonus points, e-coupons, e-miles etc. Typically the value-points are distributed by merchants, not purchased by consumers. The most ambitious schemes were Beenz and Flooz, which allowed consumers to earn Beenz/Flooz for performing activities such as visiting a website or shopping online. These units could then be spent online with participating merchants. Although representing their own unit of account they were eventually tied to each other and to the dollar (1 USD=1 Flooz=200 Beenz).[6]

5. *(Micro)Billing.* In these schemes an intermediary, typically an Internet Service Provider (ISP) or Telco, aggregates payments before they are billed. In general these use an existing billing relationship. For example, small Internet purchases are added to the ISP bill, ring tones are charged to the mobile phone account, and visits to adult entertainment sites are charged to the fixed phone-line account using the 900 system. For example iBill lets users dial a 900 number to obtain a password for a website.

6. *Mobile phone payment schemes* like Paybox (Germany) and MovilPago (Spain). These systems were designed to enable payments by mobile phone either in physical outlets or to purchase goods or services by phone.

[6] "E-currencies are going offline" TechWeb Finance, December 13, 2000. www.techweb.com/wire/story/INV20001213S0003, accessed May 5, 2003.

They generally involved establishing an account that could be filled by transferring money from a bank account. MovilPago was initially targeted at teenagers enabling them to pay in discotheques and bars.

7. *Unsafe use of offline instruments.* Böhle and Krueger do not include the 'low tech' use of traditional instruments like giving a credit card number over the phone or Internet (with minimal or no protection), or paying Internet orders by an ordinary credit transfer instruction. However, given the prevalence of their usage, I consider them as a separate category.

The framework of Böhle and Krueger is reproduced in table 9.1 with the number of schemes mentioned by them in each category and geography.[7] For example, Böhle and Krueger counted no less than 24 European schemes for making mobile payments. They count twice as many initiatives in Europe as in the US, even though the Internet boom was largely led by the US. This is of course caused by the fact that often each European country had its own initiative(s); this lends further support to the claim that payment schemes are national. I have added a column with the successful schemes (defined as performing over 10 million transactions in 2002), as well as the main failures for those categories without successes. The use of traditional instruments without any feature like SET is not included in the figures in the table.

9.1.2 Taking stock anno 2003

The only new schemes to be used for Internet purchases in any volume are PayPal and several Micro-billing schemes. All the other schemes have either gone bankrupt, refocused on other activities, or have a marginal existence. Let me give a short review per category.

1. Enhanced access to offline payment instruments. None of the high-tech access products made it: I-Pay has been withdrawn by Interpay, and SET has been abandoned by most banks. No bank wants to be seen as killing an agreed upon standard, but most of the banks are not committing resources to SET introduction:. "SET is dead".

2. E-wallets/Virtual accounts. E-wallets have sofar been a failure, with the exception of PayPal which I will discuss below. The most prominent, CyberCash launched pilots but never got critical mass. It was purchased in 2001 by

[7] There is some duplication since about 5 schemes (like CyberCash and Paypal) figure in both the US and Europe. However, the authors counted schemes with a presence in multiple European countries as a single scheme. Their numbers include schemes that were already defunct at the time as well as schemes that were announced, but not yet operational, when their paper was written (spring 2001).

TABLE 9.1 Electronic Internet payment methods in EU and US

		Eur	US	Successes	Failures
Safe use of:	-Transfers	2	1		
	-Electr. Cheques		2		BillPoint
	-Debit Cards	11	3		I-Pay
	-Credit Cards	8	5		SET
E-wallets		10	11	PayPal	CyberCash
Prepaid:	-Single purpose	1	1		
	-E-purse	17	2	Proton, Chipknip	Mondex
	-Netw. money	5	1		eCash
	-Dedicated	14	13		VisaBuxx
Money surrogates		9	11		Beenz
(Micro)Billing		12	3	ISPs, mobile operators	
Mobile payment systems		24	3		PayBox
Unsafe use of offline instr.		n.a.	n.a.	Cr. cards, transf.	Checks
Total		113	56		

Source: Böhle and Krueger (2001).

VeriSign and FirstData and switched its business to supporting encryption in online banking software.[8]

3. Prepaid products. Prepaid accounts are being used, but not for remote purchases. Single purpose schemes are doing well especially in public transportation. As for e-purses: Van Hove, 2000, provides a numerical evaluation of how they fared. The high profile trials in Atlanta, Swindon and Upper West Side were not very successful. Consumer and merchants did not see the e-purses as a significant improvement over cash and debit cards. Mondex was acquired by MasterCard, but failed both as an e-purse and as network money. Similarly, the large scale roll-outs of e-purses in the Benelux had a difficult start but recently Proton and Chipknip seem to enjoying some growth, especially in unattended points of sale like vending machines and parking meters; however, e-purses are not used for Phone or Internet payments.[9] Network money did not gain critical mass. DigiCash was bought by eCash in 1999, which in turn was acquired by Infospace in 2002, and now focuses on "comprehensive payment solutions for businesses".[10]

[8] "Acquisition of CyberCash assets will add 20,000 more merchants" (www.verisign.com/corporate /new/2001/pr_20010423.html, accessed May 5, 2003).

[9] Van Hove, 2002.

[10] "e-Gold: press release re. DigiCash asset sale"; August 1999 (www.eros-or.org/pipermail /e-lang/1999-August/002701.html, accessed May 5, 2003), and E-cash press release of Feb. 19th, 2002 (www.ecash.net/info.ecash /?ran=4871, accessed May 5th, 2003).

4. Money surrogates. Beenz and Flooz went out of business within a week of each other in August 2001, and the other money surrogates are either gone or on life-support.[11]

5. Micro-billing schemes. These *are* being used for both Internet (with payments for services being charged to the ISP account or added to the telephone bill through passwords obtained through '900' phone numbers) and mobile phone payments.

6. Mobile phone schemes. None of the schemes in this category survives; shareholder Deutsche bank sold Paybox to its management, who repositioned as a solution provider for business to business purposes.[12]

7. Unsafe use of offline instruments. The real winner, at least in the US, is the 'bare' credit card without wallets or SET. It is used for 82% of total e-commerce, with ACH and offline billing (cash on delivery and/or the buyer sends a check) making up the remainder.[13] In Europe the UK shows the same pattern as the US, but credit cards represent only 20% of Internet payments in Germany and the Nordic countries.[14] In Germany most transactions are paid through direct debit, whereby the purchaser entitles the merchant to debit his bank account.

To conclude this paragraph, let me say a word about PayPal, the only real success among the new schemes. PayPal started with 24 users in October 1999. Using its clever form of 'viral marketing' PayPal quickly gained mass among users of the eBay auction system. In July 2002 eBay bought PayPal in a $1.3 billion stock deal, announcing it would phase out Billpoint, eBay's own payment platform. At that time PayPal had 15.4 million account holders and a transfer volume of $ 1.6 billion per quarter.[15] PayPal uses the credit card system to make actual payments, and one if its main attractions has been that it allows even small merchants to accept credit card payments, a practice that has resulted in constant tensions with the credit card networks.[16]

---

[11] "Beenz.com closes Internet currency business", E-commerce times, August 17th, 2001, (www.ecommercetimes.com /perl/story/12892.html, accessed April 28th, 2003).

[12] www.paybox.de/3416.html, accessed on 4/28/03.

[13] "Statistics for electronic transactions", (www.epaynews.com/statistics/ transactions.html, accessed August 19th, 2002).

[14] BCG (2000), p18.

[15] "eBay buys PayPall", Associated Press, July 8th, 2002, (www.paypal.com/html/press /070802APEBays.html, accessed May 5th, 2003).

[16] "PayPal faces more lawsuits, could lose MasterCard". www.ecommercetimes.com/perl/story/17787.html, accessed May 21, 2003.

9.2 Existing literature and theory on electronic money and m-/e-payments

Literature and theory on these systems come from several angles. Many central banks were concerned with the monetary consequences of e-cash. Would it lead to a truly private currency and money? Would it replace the currency of central banks and even the money created by private banks? What would this mean for seigniorage revenues of central banks and their ability to regulate money supply? Much of this literature was already reviewed in chapter 2. Almost all authors agree that the monetary impact of e-cash will be limited, as cash is mainly used for illegal purposes and/or hoarding.

Several authors focus on e-purses and try to explain their (lack of) success. Van Hove (2002) blames the fact that much of the true cost of cash is hidden from consumers.[17] Chakravorti (2000) argues that e-purses so far have failed to deliver sufficient benefits to consumers compared to the alternatives. Westland (2002) surveys participants in the Hong Kong Mondex trial and concludes that the product is insufficiently differentiated from alternatives (especially debit cards) to overcome switching costs for both merchants and consumers. The observations of both authors are confirmed by the fact that e-purses are now primarily used at unmanned POS like parking meters and vending machines where the use of cash is inconvenient and debit or credit cards are not accepted.

Similar questions can be asked about Internet payment schemes. MacKie-Mason and White (1996) score 10 US schemes against 30 characteristics, such as being divisible, easily exchangeable, non-refutable, etc. They conclude that a user-centric approach may lead to the use of more than one payment mechanism. Walczuch and Duppen (2000) analyze survey data, and find that security, reliability and privacy are the most important features of a payment system for Internet purchases. They acknowledge that actual payment behavior is very different (using 'unsafe' credit cards), and conclude that the fact that people have no safe mechanisms at their disposal is one of the main reasons that online sales have not lived up to expectations (which seems a rather bold conclusion given their facts). Their results are also at odds with those found in ABA (2001); this survey among US consumers finds that security ranks only fourth

[17] Indeed several studies by De Grauwe, Buyst, et al. (2000), Jaarsma and van Rijt-Veltman (2000) find the social cost of a cash transaction to be substantial, and higher than the cost of a card transaction for transactions over EUR 13. Interestingly, Lacker (1996) argues exactly the opposite. Much of the cost of cash, the opportunity cost in the form of lost interest, is actually not a social cost since it is income for central banks (seigniorage). However, many e-cash schemes replace cash with schemes that do have real social cost, hence he argues that from a social welfare perspective, e-cash is more expensive than cash.

among consumer criteria for selecting an Internet payment instrument, after convenience, speed and ease of use.

The only online application where credit cards are not practical is what McHugh (2002) refers to as online person-to-person (P2P) payments. This is because accepting credit cards requires a merchant contract with an acquiring bank, which is not available for private individuals and cumbersome for very small businesses. McHugh quotes an estimate of Tower Group putting the US online P2P market at 100 million transactions in 2001, growing to 4 billion in 2005. The vast majority of these (55%-95% depending on the source) take place through online auctions.[18] The article compares 7 payment schemes that specifically target the online P2P market, of which PayPal has been the only real success, and even PayPal's success is "limited mainly to closed-loop environments".[19] McHugh attributes this relative success (compared to electronic money) to the fact that the online P2P schemes "leveraged past payment innovations and existing networks rather than building entirely new ones".

Finally, it is perhaps interesting to hear two entrepreneurs perform a 'post-mortem' on their company. Steve Crocker, one of the founders of CyberCash, blames: (1) too much focus on security which made the product too complicated; (2) too few merchants accepting the wallet; and (3) the lack of a real market for micro payments on the Internet: the advertising and subscription business models appeared to work better.[20] In his speech titled "How not to start a payment network", Charles Cohen, the CEO of Beenz, blames the fact that the cost of cash is a given, and accepted by society.[21] However, the cost of new schemes come on top of the cost of cash, and people are unwilling to pay for these new instruments because there are insufficient extra benefits.

Overall the literature offers explanations that appear to fit the model developed in the previous chapter: new technologies require incremental profits if they are to be adopted. This incremental profit is the difference between incremental benefits *and* costs. Technologies that leverage the existing infrastructure may well have an advantage, even if other technologies offer higher benefits. In a way, this is the same mechanism that is at work in adoption of upgrades (technology F) in the model of chapter 3. Proposition 3.2 showed

---

[18] Celent puts it at 55%, Tower Group at 95%; quoted by McHugh (2002).

[19] McHugh (2002). Since his article was published, most of the other 6 have not done well. BillPoint and eMoneyMail have been dropped by their owners (eBay and Bank One). C2it and MoneyZap have been redirected (by Citibank and WesternUnion/FirstData) towards the international P2P market, mainly remittances by foreign workers. Certapay was acquired by 5 Canadian banks that have confined its ambitions to Canada. The last one, PayDirect of Yahoo/HSBC is still active in the Yahoo auction circuit, but rumored to be unsuccessful.

[20] Crocker (1999).

[21] Cohen (2002).

how firms may adopt an upgrade F to technology f, even if F offers less benefits than another network technology g. The model of chapter 8 generalizes this mechanism to a context of multiple technologies and multiple infrastructure components.

The explanations that are offered in the literature thus seem correct and in line with model. The ambition of this chapter is therefore not to offer an alternative explanation, put rather to put the explanations for the failure of individual schemes in a common overall framework.

9.3 Applying the model

The technology succession model described in the previous chapter considers technologies as a combination of required infrastructure components and benefit elements delivered. Technologies are adopted if the additional benefits delivered are greater than the costs of the additional infrastructure required. This implies that successful technologies are those that provide extra (valuable) benefits, but do so largely on the back of existing infrastructure.

To formulate the new payment technologies in these terms, we need to select relevant infrastructure components and benefit elements. For each technology we then need to determine which of these cost and benefits components are additional to the installed base of payment technologies. To keep matters simple I propose to use just three cost/infrastructure components: the unit of account, consumer components and merchant components. The unit of account consists of the frame of reference of people in terms of prices. A new currency may either use the existing unit of account (USD, EUR) or introduce its own, as most surrogate monies did.[22] Consumer components include everything at the consumer end: accounts, software on the PC, hardware like smart card readers etc. Similarly the merchant components include: accounts (including an 'acceptance contract' with a bank or other), hardware, software, etc.

Table 9.2 compares the new schemes described by Böhle and Krueger (2001) in terms of these three components. Since schemes in a category tend to be similar in terms of usage of infrastructure, the comparison is done at the category level; where only part of the schemes in a category leverage existing infrastructure, these schemes are indicated in parentheses. The last column indicates whether the category was successful.

In reviewing the table, there appears to be a strong relation between success and the usage of existing consumer components. Basically only schemes which

[22] For a discussion of the relevance of the unit of account, see Schmitz (2001).

TABLE 9.2 Usage of existing infrastructure by new payment schemes

		Leverage existing infrastr.			Succesful
		Unit of acc	Consumers	Merchant	
Safe use of:	-Transfers	√			-
	-Electr. Chks	√			
	-Debit Cards	√			
	-Credit Cards	√			-
E-wallets		√		√(PayPal)	√(Paypal)
Prepaid:	-Sinlge purp.				-
	-E-purse	√	√		√(U-POS)
	-Netw. money	√			-
	-Dedicated	√		√(VisaBuxx)	-
Money surrogates					
(Micro)Billing		√	√		√
Mobile payment systems		√			-
Offl. instr:	-Credit Cards	√	√(US,UK)		√
	-Transfers	√	√(GE)	√	√

Note: U-POS means unmanned Point of Sale.

did not require the consumer to put in place new accounts, software etc. were successful. The exception is PayPal which does require a consumer to set up an account.

It follows that the additional benefits of all these schemes over the basic alternatives of unprotected credit cards or offline payment may be insufficient to overcome the extra cost of installing the consumer infrastructure. Again the exception is PayPal, where the additional benefit of being able to receive payments for consumers or small merchants was sufficient to get consumers to adopt new infrastructure; however (1) the effort required of consumers is small and leverages the existing E-mail and credit card account of a consumer, and (2) even with this small effort PayPal may find it difficult to grow beyond the online auction network of eBay, where the ability of small merchants to receive payment is an absolute necessity.

PayPal is therefore a good illustration the mechanism behind my model. It's success is based on the fact that PayPal provides a real benefit: the ability for small merchants and individuals to receive payments; even the unprotected use of credit cards cannot provide this benefit. At the same time PayPal requires very little infrastructure. Almost all other schemes did not enable something sufficiently new that could not be done with an unprotected credit card or transfer, while these schemes all required additional infrastructure.

9.4 Conclusions of Internet payment case

Overall the case confirms how difficult it is to introduce new payment mechanisms. It appears that the network effects inherent in payment systems play a major role. Those instruments that leverage an already installed base made it. And the only exception to this rule, PayPal established itself in a sub-network of users that interact intensively with each other (eBay). The case also confirms how past differences continue themselves after the arrival of new technologies, due to the installed base effect. The credit card countries (US and UK) use credit cards on the Internet, while a giro country like Germany relies on direct debit or other offline payment methods.

Therefore two results of the model in chapter 8 apply to the adoption of new payment instruments for Internet and mobile phone. Path dependence did occur; Germany and the US did adopt different solutions due to their different installed base. And global economies of scale did not eliminate this lack of convergence.

10
Conclusions

Chapter 1 sketched the payment instrument landscape. It provided evidence of significant and persistent differences between countries in the use of payment instruments. In addition, it found differences in the succession of payment instruments. US and Netherlands, for example, appear to have followed different technology paths in their adoption of payment instruments over the past 80 years. Based on these observations, the final section of chapter 1 formulated four central questions to be addressed in this thesis: (1) why did some countries adopt ACH/giro, while others did not? (2) why do these differences persist? (3) why do countries follow different paths? and (4) how are these differences likely to evolve in the future?

Chapters 3 and 4 introduced theoretical models to answer the first pair of questions. Using these models I obtained the following results.

a. The adoption of unsponsored network technologies requires a critical mass of adopters equal to the cost/benefit ratio of the technology. As a result, lock-in into an older and economically inferior technology may occur if the market is fragmented.

b. The existence of semi-autarkic groups (countries) that transact mostly internally, fosters the local adoption of a new network technology but also introduces another potential problem: the adoption of multiple incompatible standards or technologies by various autarkic groups.

c. Sponsored standards allow the adopting firms to appropriate (part of) the network externality. But sponsoring does not reduce the parameter range where lock-in into an inferior standard can occur, as long as the externality is relatively small (as defined by the DePalma-Leruth condition) and demand is inelastic ($\varepsilon < 1$).

d. In a world of national autarkic transaction patterns, standards will tend to be national. There is one exception: a large player (more than half the market) can keep a sponsored standard with a very low cost benefit ratio (less than 30-40%) proprietary and deny access to the other players. But even then, there will not be two competing standards; instead there will

be a standard used by the large player, while the others will use the old technology.

Chapter 8 developed a model for technology succession to analyze questions 3 and 4. This model yielded the following results:

e. The installed base of technologies strongly influences the adoption of new network technologies. Differences in the installed base often lead to differences in the adoption of new technologies, even if all countries have access to the same new technologies. This effect is strong enough to overcome significant global scale economies.

f. In the face of rapidly arriving new network technologies a late start is, on average, an advantage: it allows a country to select from a wider array of potential technologies, unhampered by an installed base. The hampering effect of the installed base is generally more important than its role as a stepping stone for new technologies.

These outcomes sketch a picture of national standards and technological trajectories. Initial differences occur because of either differences in industry structure, timing (some country have a late start) or other coincidental events. These initial differences then perpetuate themselves, turning the national standards into national paths. As stated in the preface, theoretical modelling was only one of three pillars of the reasoning in this thesis. A second pillar is formed by existing empirical evidence (reviewed in chapters 2 and 7) that most, or even all, new payment instruments are subject to network effects. This means that the models of chapters 3 and 4 can be applied to payment instruments, provided that key parameters are in the range where the outcomes are valid: inelastic demand ($\varepsilon < 1$), sufficient differentiation ($b < t$) and high autarky ($\delta < 1$). One of the results of these models in chapters 3 and 4 is that in equilibrium, standards tend to be national. This in turn is one of the main assumptions of the model in chapter 8. The third pillar, the two cases described in chapters 5 and 6 provides evidence that the parameters are indeed in this safe zone. And the actual events of the three cases (the two cases of chapter 5 and 6 and the 'illustration' of chapter 9) are in line with the model predictions. This allows me to now answer the four central questions of this dissertation.

1. *Why did some countries adopt ACH/giro, while others did not?* This is caused by a combination of small historical coincidences, in particular the introduction of giro-systems, and the relatively concentrated nature of European banking in the 1960s, right before the take-off of mass consumer banking. The fact that most of continental Europe came somewhat later

to mass consumer banking (in the 1960s as opposed to the 1950s for the US) may also have played a role. By then high volume processing techniques were becoming available to facilitate a giro-system.

2. *Why do these differences persist?* The US continues to use checks because the shift from checks to ACH requires concerted action by a very large group of banks. Subsequent upgrades of check technology have closed the gap with ACH/giro somewhat (e.g. check truncation and verification at the POS); but this has further raised the critical mass that is needed for ACH adoption. European countries continue to use different ACH systems because of highly autarkic transaction patterns: over 99% of all transfers take place *within* the same country. This makes migration to a common standard insufficiently attractive. The adoption of an overlay for cross-border transactions seems more likely for the near to medium future.

3. *Why do countries follow different paths?* Because the adoption of payment technologies is greatly facilitated if existing infrastructure elements can be used. As a result, differences in the installed base lead to differences in the adoption of new payment instruments.

4. *How are these differences likely to evolve in the future?* The differences are likely to persist for quite a while. Even the advent of radically new technologies or technology elements, like the Internet, may not erase the differences. Internet payments appear to follow pre-existing national differences in instruments: credit cards in the US and UK, direct debit and transfers in Germany.

11
Epilogue: "To a man with a hammer ..

..an increasing number of problems look like nails".[1] Having completed the above argumentation I am a man with a hammer. As argued in the preface, careless application of the concept of network effects usage can lead to serious injury. Nevertheless, in this epilogue, let me indulge in temptation and suggest some nails to be hit.

The mechanism of deeply rooted local standards that are very persistent, even in the face of new global developments, may be at work in other places where country or regional differences continue to exist. As was discussed in chapter 2, several authors like Paul Krugman and Jean-Michel Dalle have already applied local externalities to model technology adoption.[2] The next candidate could be capital. Prescott forcefully argues in his 1997 Klein lecture (Prescott 1998), that differences in measured capital just cannot explain productivity differences between countries. And the measurement of capital in modern economics seems hopelessly out of touch with reality. It still measures things like machines and buildings where most businesses today would consider brands, customer bases, established procedures and systems, and trained employees their true capital. Indeed, one may argue that standards are a form of capital for firms and countries. Microsoft's true capital is the fact that it owns the desktop standard. The fact that European countries have an installed and accepted base of ACH systems enables them to achieve a much higher productivity in transfer payments. At the same time this installed base may make it more difficult to adopt new payment technologies.

One could argue that many standards like a legal system, social and business codes of conduct, etc. are a form of capital. Applying my model to this notion of capital leads to some interesting implications. For example, more capital is not always better.[3] And changing the installed base is not trivial: reaching another peak may involve retracing steps (undoing earlier adoptions) to get

---

[1] Anonymous.

[2] Krugman (1994), Dalle (1997).

[3] And indeed Durlauf (1999), in his "case against social capital", argues that social capital (for example established communities) can have both good and bad effects, where most advocates associate it exclusively with positive effects.

to a different path. As a result imitating the leader is not necessarily the best way to raise the productivity of a country. The right solution for one country may well be the wrong solution for another country with a different installed base.

To conclude, it may well be that we have only scratched the surface of network effects and their implications.

References

ABA (2001). Study of consumer payment preferences, American Bankers Association with Dove Consulting.

Abernathy, W. J. and J. M. Utterback (1975). A dynamic model of process and product innovation by firms, Center for Policy Alternatives, CPA 75-6.

Abramovitz, M. (1986). "Catching up, forging ahead, and falling behind." Journal of Economic History 46(2): 385-406.

Ackerberg, D. A. and G. Gowrisankaran (2002). Quantifying equilibrium network externalities in the ACH banking industry. Mimeo, UCLA.

Alessie, R., R. Gradus, et al. (1990). "The problem of not observing small expenditures in a consumer expenditure survey." Journal of Applied Econometrics 5: 151-166.

Amable, B. and R. Boyer (1995). "Europe in the world technology competition." Structural Change and Economic Dynamics 6: 167-183.

Anderson, S. P., A. DePalma, et al. (1992). Discrete choice theory of product differentiation. Cambridge, MIT Press.

Arthur, B. W. (1989). "Competing technologies, increasing returns, and lock-in by historical events." Economic Journal 99: 116-131.

Atkinson, A. B. and J. E. Stiglitz (1969). "A new view of technological change." Economic Journal 79(315): 573-578.

Ausubel, L. M. (1991). "The failure of competition in the credit card market." American Economic Review 81/1: 50-81.

Avery, R. B., G. E. Elliehausen, et al. (1986). "The use of cash and transaction accounts by American families." Federal Reserve Bulletin Feb. 1986: 87-108.

Bak, P., S. F. Nørrelykke, et al. (1999). "Dynamics of money." Physical Review 60(3): 2528-2532.

Balto, D. A. (1995). "Payment systems and antitrust: can the opportunities for network competition be recognized?" Federal Reserve Bank of St. Louis, Review 77/6: 19-40.

Balto, D. A. (2000). "The problem of interchange fees: costs without benefits?" European Competition Law Review 21/4: 215-223.

Bassanini, A. P. and G. Dosi (1998). Competing technologies, international diffusion and the rate of convergence to a stable market structure. IIASA interim report, no. IR-98-012.

Bauer, P. W. and G. D. Ferrier (1996). "Scale economies, cost efficiencies, and technological change in federal reserve payments processing." Journal of Money, Credit and Banking 28/4: 1004-1039.

Baumol, W. J. (1952). "The transactions demand for cash: an inventory theoretic approach." Quarterly Journal of Economics 66: 545-556.

Baxter, W. (1983). "Bank interchange of transactional paper: legal and economic perspectives." Journal of Law and Economics 26: 541-588.

BCG (2000). The race for online riches, Boston Consulting Group, www.dad.be/library/pdf/BCG1.pdf, accessed May 21, 2003.

Ben-David, D. (1993). "Equalizing exchange: trade liberalization and income convergence." Quarterly Journal of Economics 108(3): 653-679.

Besen, S. M. and J. Farrell (1994). "Choosing how to compete: strategies and tactics in standardization." Journal of Economic Perspectives 2-Aug: 117-131.

BIS (2003 and previous years). Statistics on payment and settlement systems in selected countries (red book). Basel, Bank for International Settlements.

Boeschoten, W. C. (1992). Currency use and payment patterns. PhD dissertation, University of Amsterdam.

Boeschoten, W. C. (1998). "Cash management, payment patterns and the demand for money." De Economist 146/1: 117-142.

Boeschoten, W. C. and M. M. G. Fase (1989). "The way we pay with money." Journal of Business & Economic Statistics 1989/7: 319-326.

Boeschoten, W. C. and M. M. G. Fase (1992). "The demand for large bank notes." Journal of Money, Credit and Banking 24/3: 319-337.

Boeschoten, W. and G. E. Hebbink (1996). Electronic money, currency demand and seigniorage loss in the G-10 countries. De Nederlandsche Bank Staff Report No. 1, 1996.

Böhle, K. and M. Krueger (2001). Payment culture matters - a comparative EU-US perspective on Internet payments -. Background Paper No. 4, Electronic Payment Systems Observatory (ePSO), Report EUR 19936 EN.

Bresnahan (2001). Network effects and Microsoft. Stanford University, www .stanford.edu/~tbres/Microsoft, accessed 15/5/03.

Breshanan, T. F. and M. Trajtenberg (1995). "General purpose technologies: 'engines of growth'?" Journal of Econometrics 65: 83-108.

Caskey, J. P. and G. H. Sellon (1994). "Is the debit card revolution finally here?" Federal Reserve Bank of Kansas City, Economic Review: 80-95.

Caskey, J. P. and S. St. Laurent (1994). "The Susan B. Anthony dollar and the theory of coin/note substitution." Journal of Money, Credit and Banking 26/3: 495-510.

Chakravorti, S. (2000). Why has stored value not caught on?, Emerging Issues Series, Federal Reserve Bank of Chicago. S&R-2000-6.

Chakravorti, S. and T. McHugh (2002). "Why do we use so many checks?" Economic Perspectives, Federal Reserve bank of Chicago 2002(Q3): 44-59.

Chakravorti, S. and A. Shah (2001). A study of the interrelated bilateral transactions in credit card networks, Federal Reserve Bank of Chicago, Emerging Payments Occasional Paper Series (EPS-2001-2).

Chang, H. H. and D. Evans (2000). "The competitive effects of the collective setting of interchange fees by payment card systems." Antitrust Bulletin: 641-677.

Chaum, D. (1992). "Achieving electronic privacy." Scientific American(August 2002): 96-101.

Choi, J. P. (1996). "Standardization and experimentation: ex ante vs. ex post standardization." European Journal of Political Economy 12: 273-290.

Church, J., N. Gandal, et al. (2002). Indirect network effects and adoption externalities, CEPR discussion paper no.3738.

Church, J. and I. King (1993). "Bilingualism and network externalities." Canadian Journal of Economics 26/2: 337-345.

Cohen, C. (2002). "How not to start a payment network". Presentation to Digital money forum 2002.

Costa, C. and P. De Grauwe (2001). Monetary policy in a cashless society. SUERF meeting, Brussels.

Cowan, R. (1990). "Nuclear power reactors: a study in technological lock-in." Journal of Economic History 50/3: 541-567.

Cowan, R. (1991). "Tortoises and hares: choice among technologies of unknown merit." Economic Journal 101: 801-814.

Cowan, R. and W. Cowan (1998). Technological standardization with and without borders in an interacting agents model. Merit research memorandum # 2/98-018.

Cowan, R. and D. Foray (2000). On the nature and role of counterfactual history as an empirical tool in economics. Mimeo, Merit University of Maastricht.

Cowan, R. and P. Gunby (1996). "Sprayed to death: path dependence, lock-in and pest control strategies." Economic Journal 106: 521-542.

Cowan, R. and S. Hultén (1996). "Escaping lock-in: the case of the electric vehicle." Technological Forecasting and Social Change 53: 61-79.

Crocker, S. (1999). The vast (?) market for micro payments. Information Impacts Magazine. www.cisp.org/imp-/April_99/04_99crocker.htm, accessed on 5/5/03.

Dalle, J.-M. (1997). "Heterogeneity vs. externalities in technological competition: a tale of possible landscapes." Journal of Evolutionary Economics 7: 395-413.

Daniels, K. N. and N. B. Murphy (1994). "The impact of technological change on the currency behavior of households: an empirical cross-section study." Journal of Money, Credit and Banking 26/4: 867-874.

David, P. A. (1985). "Clio and the economics of QWERTY." Economic History 75/2: 333-337.

David, P. A. (1994). "Why are institutions the "carriers of history"?: path dependence and the evolution of conventions, organizations and institutions." Structural Change and Economic Dynamics 2/5: 205-220.

David, P. A. and D. Foray (1994). Percolation structures, Markov random fields and the economics of EDI standards diffusion. Global telecommunications strategies and technological changes. G. Pogorel (ed.), Elsevier Science B.V.

David, P. A., D. Foray, et al. (1998). "Marshallian externalities and the emergence of spatial stability of technology enclaves." Economics of Innovation and New Technology 6: 147-182.

David, P. A. and Greenstein (1990). "The economics of compatibility standards: an introduction to recent research." Economics of Innovation and New Technology 1: 3-44.

De Grauwe, P., E. Buyst, et al. (2000). The cost of cash and cards compared. The cases of Iceland and Belgium. Mimeo, University of Leuven.

DeGreyse, H. and S. Ongena (2002). Distance, lending relationships and competition, Centre for studies in economics and finance, Fisciano, working paper number 80.

DePalma, A. and L. Leruth (1993). "Equilibrium in competing networks with differentiated products." Transportation Science 27/1: 73-80.

Dosi, G. (1982). "Technological paradigms and technological trajectories." Research Policy 11: 147-162.

Dosi, G. (1997). "Opportunities, incentives and the collective patterns of technological change." Economic Journal 107: 1530-1547.

Dosi, G., C. Freeman, et al. (1988). Technical change and economic theory. London, Printer publishers.

Dowd, K. and D. Greenaway (1993). "Currency competition, network externalities and switching costs: towards and alternative view of optimum currency areas." Economic Journal 103: 1180-1189.

Dranove, D. and N. Gandal (1999). Network effects, standardization, and the Internet: what have we learned from the DVD vs. DIVX battle? CEPR Discussion paper 2335.

Drehmann, M., C. Goodhart, et al. (2001). "The challenges facing currency usage: will the traditional transaction medium be able to resist competition from the new technologies." Economic Policy 34: 195-227.

Duca, J. V. and W. C. Whitesell (1995). "Credit cards and money demand." Journal of Money, Credit and Banking 27/2: 604-623.

Durlauf, S. N. (1999). The case against social capital. University of Wisconsin, www.ssc.wisc.edu/econ/archive/wp9929.pdf, accessed 15/5/03.

ECBS (2003). Summary of European account numbers, ECBS TR 201V2.2.22. Brussels, European Committee for Banking Standards.

Economides, N. and F. Flyer (1998). "Equilibrium coalition structures in markets for network goods." Annales d'Economie et de Statistique 49/50: 361-380.

Economides, N. and C. Himmelberg (1995). Critical mass and network size with application to the US fax market, Discussion paper EC-95-11, Stern School of Business, NYU.

Elliehausen, G. E. and J. D. Wolken (1992). "Banking markets and the use of financial services by households." Federal Reserve Bulletin 78: 169-181.

Ellison, G. (1993). "Learning, local interaction, and coordination." Econometrica 61/5: 1047-1071.

European Banking Federation (2002a). European Banks launch "Euroland- our Single Payments Area". Press release. Brussels.

European Banking Federation, European Savings Banks Group, et al. (2002b). "Euroland: our single payment area!". White paper. Brussels.

European Commission (1997). Cross border credit transfers directive. Directive, 97/5/EC.

European Commission (2000). Commission plans to clear certain Visa provisions, challenge others. Press release IP/00/1164.

European Commission (2001). Study on the verification of a common and coherent application of directive 97/5/EC on cross-border credit transfers in the 15 member states. Retail Banking Research, London.

European Community (2001). Regulation on cross-border payments in Euro. Regulation no. 2560/2001.

Evans, D. and R. Schmalensee (1999). Paying with plastic, the digital revolution in buying and borrowing, MIT Press.

Farrell, J. and G. Saloner (1985). "Standardization, compatibility and Innovation." RAND Journal of Economics 16: 70-83.

Farrell, J. and G. Saloner (1986). "Installed base and compatibility: innovation, product preannouncements, and predation." American Economic Review 76/5: 940-955.

Farrell, J. and G. Saloner (1988). "Coordination through committees and markets." RAND Journal of Economics 19/2: 235-252.

Fase, M. M. F. (1999). Geld in het fin de siècle, Amsterdam University Press.

Flatraaker, D. and P. Robinson (1995). "Income, costs and pricing in the payment system." Norges Bank Economic Bulletin 66: 321-332.

Foray, D. (1997). "The dynamic implications of increasing returns: technological change and path dependent efficiency." International Journal of Industrial Organization 15: 733-752.

Frankel, A. B. and J. C. Marquardt (1987). International payments and EFT links. Electronic funds transfers and payments: the public policy issues. E. H. Solomon. Dordrecht, Kluwer-Nijhoff Publishing: 111-130.

Freeman, C. and C. Perez (1988). Structural crises of adjustment, business cycles and investment behaviour. Technical change and economic theory. G. Dosi, C. Freeman, R. Nelson, G.Silverberg and L. Soete (eds.). London, Printer Publishers.

FRS (2002). Retail Payments Research Project; a snapshot of the U.S. payments landscape. Washington D.C., Federal Reserve System.

FRBSF (Federal Reserve Bank of San Francisco) (2002). "Trends in the concentration of bank deposits." FRBSF Economic Letter 21(July).

Gandal, N. (1994). "Hedonic price indexes for spreadsheets and network externalities." RAND Journal of Economics 25/1: 160-170.

Gans, J. S. and S. P. King (2003). "Approaches to regulating interchange fees in payment systems." Review of Network Economics 2/2: 125-145.

Gerdes, G. R. and J. K. Walton II (2002). "The use of checks and other non-cash payment instruments in the United States." Federal Reserve Bulletin(August 2002): 360-374.

Glaser, P. F. (1988). Using technology for competitive advantage: the ATM experience at Citicorp. Managing innovation: cases from the financial services industries. B.R. Guile and J.B. Quinn (eds.). Washington D.C., National Academy Press: 108-114.

Gong, Y. and W. Jang (1998). "Culture and development: reassessing cultural explanations on Asian economic development."

Good, B. A. (1997). Electronic money, Federal Reserve Bank of Cleveland working paper #9716, August 1997.

Gort, M. and S. Klepper (1982). "Time paths in the diffusion of product innovations." Economic Journal 92(367): 630-653.

Gowrisankaran, G. (1999). Issues and prospects for payments systems deregulation. Working paper, www.econ.umn.edu/~gautam/-pdf_papers/paydereg.pdf, accessed May 21, 2003.

Gowrisankaran, G. and J. Stavins (2002). Network externalities and technology adoption: lessons from electronic payments. NBER Working Paper No. 8943.

Graham, I., G. Spinardi, et al. (1995). "The dynamics of EDI standards development." Technology Analysis and Strategic Management 7(1): 3-20.

Griliches, Z. and J. Schmookler (1963). "Inventing and maximizing." American Economic Review 53(4): 725-729.

Gross, D. B. and N. S. Souleles (2002). "Do liquidity constraints and interest rates matter for consumer behavior? Evidence from credit card data." Quarterly Journal of Economics 117(1): 149-185.

Gunby, P. (1996). Explaining adoption patterns of process standards. PhD dissertation, University of Western Ontario.

Hancock, D. and D. B. Humphrey (1998). "Payment transactions, instruments, and systems: a survey." Journal of Banking and Finance 21: 1573-1624.

Hannan, T. H. (2001). "Retail fees of depository institutions." Federal Reserve Bulletin 2001(January): 1-11.

Hannan, T. H. and J. M. McDowell (1984). "The determinants of technology adoption: the case of the banking firm." RAND Journal of Economics 15/3: 328-335.

Hannan, T. H. and J. M. McDowell (1987). "Rival precedence and the dynamics of technology adoption: an empirical analysis." Economica 54: 155-171.

Harley, C. K. (1973). "On the persistence of old techniques: the case of North American wooden shipbuilding." Journal of Economic History 33/2: 372-398.

Hirschman, E. C. (1982). "Consumer payment systems: the relationship of attribute structure to preference and usage." Journal of Business 55(4): 531-545.

Hogesteeger, G. and H. de Lanoy Meijer (eds.) (1992). Over automatisering gesproken : dertien gesprekken met kopstukken uit de geschiedenis van de automatisering in Nederland. Amsterdam : Commissie Historie Automatisering in Nederland (CHAN).

Hotelling, H. (1929). "Stability in competition." Economic Journal 39: 41-57.

Huck, S., H.-T. Normann, et al. (2001). Two are few and four are many: number effects in experimental oligopolies. Mimeo, Royal Holloway, London.

Humphrey, D. B. and A. N. Berger (1990). Market failure and resource use: economic incentives to use different payment mechanisms. The U.S. payment system: efficiency, risk and

the role of the Federal Reserve. D. B. Humphrey (ed.). Kluwer Academic Publishers: 45-86.

Humphrey, D., A. Kaloudis, et al. (2000). Forecasting cash use in legal and illegal activities. Mimeo, Norges Bank.

Humphrey, D. B., R. Keppler, et al. (1997). Cost recovery and pricing of payment services: theory, methods, and experience. Working paper, World Bank.

Humphrey, D. B., M. Kim, et al. (2001). "Realizing the gains from electronic payments: costs, pricing and payment choice." Journal of Money, Credit and Banking 33(2): 216-234.

Humphrey, D. B., L. B. Pulley, et al. (1996). "Cash, paper, and electronic payments: a cross-country analysis." Journal of Money, Credit and Banking 28/4: 914-939.

Humphrey, D. B., L. B. Pulley, et al. (2000). "The check is in the mail: why the US lags in the adoption of cost-saving electronic payment instruments." Journal of Financial Services Research 17(1): 17-39.

Humphrey, D. B., M. Willesson, et al. "What does it cost to make a payment?" Review of Network Economics 2/2: 159-174.

Issing, O. (1999). Hayek, currency competition and European monetary union. Hayek Memorial Lecture, Institute of Economic Affairs occasional paper 111.

Jaarsma, K. and W.V.M van Rijt-Veltman (2000). De kassa rinkelt niet voor niets, afrekenen kost ook geld. The Hague, Hoofdbedrijfschap Detailhandel.

Jaffe, A. B., M. Trajtenberg, et al. (1993). "Geographic localization of knowledge spillovers as evidenced by patent citations." Quarterly Journal of Economics 108(3): 577-598.

Jaffe, A. B. and M. Trajtenberg (1999). "International knowledge flows: evidence from patent citations." Economics of Innovation and New Technology 8: 105-136.

Jonard, N. and E. Schenk (1999). A duopoly Logit model with price competition and strategic compatibility. Merit research memorandum no. 2/99-011.

Jongepier, P. (2002). "Gebruik chipkaarten neemt toe." NVB Bulletin 2002(1): 1-3.

Katz, M. L. and C. Shapiro (1985). "Network externalities, competition, and compatibility." American Economic Review 75/3: 424-440.

Katz, M. L. and C. Shapiro (1986). "Technology adoption in the presence of network externalities." Journal of Political Economy 94: 822-841.

Kauffman, R. J. and Y.-M. Wang (1994). "An exploratory econometric analysis of shared electronic banking network adoption." Journal of Strategic Information Systems 3(1): 61-76.

Kauffman, S., J. Lobo, et al. (2000). "Optimal search on a technology landscape." Journal of Economic Behavior and Organization 43(2): 141-166.

Keller, W. (2002). "Geographic localization of international technology diffusion." American Economic Review 92(1): 120-142.

Khiaonarong, T. (1997). A review of payment systems literature. Working paper Series No. 61, London School of Economics and Political Science, Department of Information Systems.

Kiyotaki, N. and R. Wright (1993). "A search-theoretic approach to monetary economics." American Economic Review 83/1: 63-77.

Kleimeier, S. and H. Sander (2000). "Regionalisation versus globalisation in European financial market integration: evidence from co-integration analyses." Journal of Banking and Finance 24: 1005-1043.

Kleimeier, S. and H. Sander (2002). Consumer credit rates in the Eurozone: evidence on the emergence of a single retail banking market. ECRI research report no. 2.

KPMG (1990). Rapportage inzake de kosten- en opbrengsten van het binnenlands betalingsverkeer over 1989 in opdracht van de Nederlandse Vereniging van Banken, Utrecht.

Krugman, P. (1994). "Complex landscapes in economic geography." American Economic Review Papers and Proceedings 84(2): 412-416.

Lacker, J. M. (1996). "Stored value cards: costly private substitutes for government currency." Federal Reserve Bank of Richmond, Economic Quarterly 82/3: 1-25.

Lafferty Group (2003). E-purses enjoy slow, steady growth. Card Payments. February 2003: 5.

Law, A. M. and W. D. Kelton (1991). Simulation modelling and analysis, McGraw-Hill.

Lelieveldt, S. L. (2000). Standardizing retail payment instruments. Information technology standards and standardization: a global perspective. K. Jacobs (ed.), Hershey: 186-197.

Liebowitz, S. J. and S. Margolis (1990). "The fable of the keys." Journal of Law and Economics 33(April 1990): 1-26.

Luce, R.D. and H. Raiffa (1957), Games and decisions, introduction and critical survey, Dover publications, New York.

Lundvall, B. A. (1988). Innovation as an interactive process: from user-producer interaction to the national system of innovation. Technical change and economic theory. G. Dosi, C. Freeman, R. Nelson, G. Silverberg and L. Soete (eds.). London, Printer Publishers.

MacKie-Mason, J. K. and K. White (1996). Evaluating and selecting digital payment mechanisms. Mimeo, University of Michigan, Ann Arbor.

Mandell, L. (1990). The credit card industry: a history, Twayne Publishers.

Mantel, B. and T. McHugh (2001). Competition and innovation in the consumer e-payments market? Considering the demand, supply, and public policy issues. Emerging Payments Occasional Working Paper Series, December 2001 (EPS-2001-4), Federal Reserve Bank of Chicago.

Massoud, N. and D. Bernhardt (2002). ""Rip-off" ATM surcharges." RAND Journal of Economics 33(1): 96-115.

Matutes, C. and A. J. Padilla (1994). "Shared ATM networks and banking competition." European Economic Review 38: 1113-1138.

Matutes, C. and P. Regibeau (1988). "'Mix and match': product compatibility without network externalities." RAND Journal of Economics 19/2: 221-234.

Matutes, C. and P. Regibeau (1992). "Compatibility and bundling of complementary goods in a duopoly." Journal of Industrial Economics 40/1: 37-54.

McAndrews, J. J. (1997). "Network issues and payment systems." Federal Reserve Bank of Philadelphia, Business Review: 15-25.

McAndrews, J. J. (2003). "Automated teller machine network pricing - a review of the literature." Review of Network Economics, 2/2: 146-158.

McAndrews, J. J. and R. Rob (1996). "Share ownership and pricing in a network switch." International Journal of Industrial Organization 14: 727-745.

McHugh, T. (2002). The growth of person-to-person electronic payments, Federal Reserve Bank of Chicago, Chicago Fed Letter no. 180.

Murphy, N. B. (1991). "Determinants of household check writing: the impacts of the use of electronic banking services and alternative pricing of services." Financial Services Review 1-Jan: 35-44.

NEI (2000). Fusies en overnames in het Nederlandse bankwezen; finaal rapport. De Nederlandse Mededingingsautoriteit. Juli 2000.

Nelson, R. R. (1998). The co-evolution of technology, industrial structure, and supporting institutions. Technology, organization and competitiveness. G. Dosi, D. J. Teece and J. Chytry (eds.), Oxford University Press.

Nelson, R. R. and S. G. Winter (1982). An evolutionary theory of economic change. Cambridge, Harvard University Press.

Nelson, R. R. and G. Wright (1992). "The rise and fall of American technological leadership." Journal of Economic Literature 30(4): 1931-1964.

NIBE Banks and Brokers in the Netherlands, years 1972 and 2002. NIBE-SVV, Amsterdam.

Norges Bank (2001). Mission impossible? Benchmarking the Norwegian payment systems using information from the BIS' Red Book tables. Mimeo, Norges Bank.

Osterberg, W. P. and J. B. Thomson (1998). Network externalities: the catch-22 of retail payments innovation. Federal Reserve Bank of Cleveland Working Paper, February 15, 1998.

Park, S. (2000). Quantitative analysis of network externalities in competing technologies: the VCR case. Mimeo, SUNY at Stony Brook.

Paroush, J. and D. Ruthenberg (1986). "Automated teller machines and the share of demand deposits in the money supply." European Economic Review 30: 1207-1215.

Patel, P. and K. Pavitt (1998). Uneven (and divergent) technological accumulation among advanced countries: evidence and a framework of explanation. Technology, organization, and competitiveness. G. Dosi, D.J. Teece, J. Chytry (eds.), Oxford University Press.

PCGD (1973). Het goud van de postgiro. Utrecht, Prisma pocket 1551, Het Spectrum.

Peekel, M. and J. W. Veluwenkamp (1984). Het girale betalingsverkeer in Nederland - uitgave ter gelegenheid van de vijf miljoenste postrekening. Amsterdam, Postgiro/Rijkspostspaarbank.

Perloff, J. M. (2001). Microeconomics, Addison-Wesley.

PIRG (1999). Big banks, bigger fees: the 1999 PIRG bank fee survey. US Public Interest Research Group, www.pirg.org/reports/-consumer/bankfees/bank99.pdf.

Plott, C. R., A. B. Sugiyama, et al. (1993). "Economies of scale, natural monopoly, and imperfect competition in an experimental market." Southern Economic Journal 59: 261-287.

Porter, M. (1990). The competitive advantage of nations. New York, Free Press.

Porter, R. D. and R. A. Judson (1996). "The location of U.S. currency: how much is abroad?" Federal Reserve Bulletin 82/10: 883-903.

Prescott, E. C. (1987). "A multiple means-of-payment model." New Approaches to Monetary Economics (book): 42-51.

Prescott, E. C. (1998). "Needed: a theory of total factor productivity." International Economic Review 39(3): 525-551.

Prescott, E. S. and J. A. Weinberg (2000). Incentives, communication, and payment instruments. Federal Reserve Bank of Richmond working paper. 00-11.

Puffert, D. J. (2001). Path dependence in spatial networks: the standardization of railway track gauge. Working paper, University of Munich.

Reserve Bank of Australia and Australian Competition and Consumer Commission (2000). Debit and credit card schemes in Australia: a study of interchange fees and access. Canberra, RBA.

Rey, P. and J. Tirole (2000). Loyalty and investment in cooperatives. Mimeo, IDEI Toulouse.

Rhoades, S. A. (2000). Bank mergers and banking structure in the United States. Washington DC, Board of Governors of the Federal Reserve System, Staff Study 174.

Rochet, J.-C. and J. Tirole (1999). Cooperation among competitors: the economics of credit card associations. Centre for Economic Policy Research, Discussion paper No. 2101.

Rohlfs, J. (1974). "A theory of interdependent demand for a communications service." Bell Journal of Economics and Management Science 5: 16-37.

Rothaermel (2000). "Technological discontinuities and the nature of competition." Technology Analysis and Strategic Management 12: 149-160.

Ruttan, V. W. (1997). "Induced innovation, evolutionary theory and path dependence: sources of technical change." Economic Journal 107: 1520-1529.

Saloner, G. and A. Shepard (1995). "Adoption of technologies with network effects: an empirical examination of automated teller machines." RAND Journal of Economics 26/3: 479-501.

Salop, S. C. (1979). "Monopolistic competition with outside goods." RAND Journal of Economics 10: 141-156.

Salop, S. C. (1990). "Deregulating self-regulated shared ATM networks." Economics of Innovation and New Technology 1: 85-96.

Santomero, A. M. and J. J. Seater (1996). "Alternative monies and the demand for media of exchange." Journal of Money, Credit and Banking 28/4: 942-960.

Schmalensee, R. (2001). Payment systems and interchange fees. Mimeo.

Schmitz, S. W. (2001). The institutional character of new electronic payments systems: the redeemability and the unit of account. Mimeo, Austrian Academy of Sciences. Version January 3, 2001.

Schmookler, J. (1966). Invention and economic growth. Cambridge, Mass., Harvard University Press.

Seitz, F. (1995). The circulation of Deutsche mark abroad. Deutsche Bundesbank, Discussion paper 1/95.

Sharma, S. (1993). Behind the diffusion curve, UCLA department of economics. Working paper #686.

Shy, O. (1996). "Technology revolutions in the presence of network externalities." International Journal of Industrial Organization 14: 785-800.

Shy, O. (2001). The economics of network industries.

Shy, O. (2002). "A quick-and-easy method for estimating switching costs." International Journal of Industrial Organization 20: 71-87.

Shy, O. and J. Tarkka (1998). The market for electronic cash cards. Bank of Finland discussion papers. 21/98.

Stalder, F. and A. Clement (1998). "Exploring policy issues of electronic cash: the Mondex case." Canadian Journal of Communication 24(2).

Stavins, J. (1997). "A comparison of social costs and benefits of paper check presentment and ECP with truncation." New England Economic Review(July/August): 27-44.

Stavins, J. (2001). "Effect of consumer characteristics on the use of payments instruments." New England Economic Review(3): 19-31.

Stone, B. K. (1990). The electronic payment industry: change barriers and success requirements from a market segments perspective. "The U.S. payment system: efficiency, risk and the role of the Federal Reserve (book)": 13-40.

Swann, P. (2002). "The functional form of network effects." Information Economics and Policy 14: 417-429.

Tak, A. and F. R. J. Dubois (1924). Rapport over de oorzaken van de ontwrichting en de verergerde ontwrichting van den postchèque- en girodienst en de schuldigen daaraan. The Hague, Algemeene Landsdrukkerij.

ten Raa, T. and V. Shestalova (2003). "Empirical evidence on payment media costs and switch points." Journal of Banking and Finance 2003(in press).

Thomson, F. P. (1964). Giro credit transfer systems. Oxford, Pergamon Press.

Tirole, J. (1989). The theory of Industrial Organization. MIT Press.

Tripsas, M. (1997). "Unraveling the process of creative destruction: complementary assets and incumbent survival in the typesetter industry." Strategic Management Journal 18: 119-142.

Tushman, M. L. and P. Anderson (1986). "Technological discontinuities and organizational environments." Administrative Science Quarterly 31: 439-465.

Van Hove, L. (2000). "Electronic purses: (which) way to go?" First Monday 5(7): (July 2000).

Van Hove, L. (2002). "Electronic money and cost based pricing." Wirschaftpolitische Blätter 2002(2).

Van Hove, L. (2003). "Cost-based pricing of payment instruments: the state of the debate." De Economist, forthcoming.

Veblen, T. (1915). Imperial Germany and the industrial revolution. New York.

Walczuch, R. and R. Duppen (2002). Payment systems for the Internet - consumer requirements. University of Maastricht.

Weinberg, J. A. (1997). "The organization of private payment networks." Federal Reserve Bank of Richmond, Economic Quarterly 83(2): 25-43.

Weinberg, J. A. (2002). "Imperfect competition and the pricing of inter-bank payment services." Federal Reserve Bank of Richmond, Economic Quarterly 88(1): 51-66.

Wells, K. E. (1996). "Are checks overused?" Federal Reserve Bank of Minneapolis, Quarterly Review 20/4: 2-12.

Westland, J. C. (2002). "Preference ordering cash, near cash, and electronic cash." Journal of organizational computing and electronic commerce 12(3): 223-242.

White, K. J. (1976). "The effect of bank credit cards on the household transactions demand for money." Journal of Money, Credit and Banking 8: 51-61.

Windrum, P. and C. Birchenhall (2000). Modeling technological successions. Mimeo, Merit University of Maastricht.

Wright, J. (2001). The determinants of optimal interchange fees in payment systems, University of Auckland.

Wolf, H. (1983). Betalen via de bank - van verleden tot heden. NIBE, serie Bank en Effectenbedrijf nr 18,. Deventer, Kluwer.

Appendix A
Proof of propositions in chapter 3

Proposition 3.2 *The availability of upgrades increases* s_c *for all combinations of* b *and* c.

Proof. Let F be an upgraded version of f. Suppose a bank can enable its customers to use technology F at a fixed cost per customer c_F and can then get benefits per transaction of b_F, with $b_F > c_F$. Assume $c_F < c$ but $b_F - c_F < b - c$: F can be obtained at a lower cost, but the ultimate profit is lower than the potential profit of adopting g. A crucial difference is that F is compatible with f, so it can always be used for *all* transactions, where g can only be used if both parties have adopted g. Depending on the parameters b, b_F, c, c_F and s_1 there may now be one or two Nash equilibria: (1) all adopt g, this is always an equilibrium, and (2) all adopt F, this is an equilibrium for some parameter values.

As before let s_g be the joint share of all banks that adopted g. All banks adopting g but not F is (still) a Nash equilibrium, since $b_F < b$ (this follows because $c_F < c$ and $b_F - c_F < b - c$) and any bank would decrease profits by adopting F in addition to g. All banks adopting F but not g can be a Nash equilibrium if for all banks i:

$$s_i(b - b_F) - c < 0.$$

i.e. adopting g on top of F is unprofitable, and

$$s_i b - c - (b_F - c_F) < 0.$$

i.e. adopting g and dropping F (reverting to f for the non-g transactions) is unprofitable. These equations reduce to:

$$s_i < \frac{c}{(b - b_F)} \text{ and } s_i < \frac{c}{b} + \frac{b_F - c_F}{b},$$

for all banks i, or

$$s_1 < \frac{c}{(b - b_F)}, \tag{A.1}$$

and

$$s_1 < \frac{c}{b} + \frac{b_F - c_F}{b}. \tag{A.2}$$

Now the right hand side of both (A.1) and (A.2) is larger than that of (3.1), so there is now a larger area where a suboptimal equilibrium occurs. Finally, all banks sticking with f is no longer equilibrium, since in that case any bank can adopt F and increase profits per customer by $b_F - c_F$, which is positive by assumption. ■

Proposition 3.4 *If there are multiple firms within semi-autarkic countries there are three equilibria:*

(a) *Adoption by all banks. This is an equilibrium for all market structures* and all $0 < \frac{c}{b} < 1$ and $0 \leq \delta \leq 1$.

(b) Adoption by no bank. This is an equilibrium if $r_{i1} < \frac{c}{b}\frac{1}{[1-\delta(1-s_i)]}$. The parameter space where this may happen increases as δ gets closer to 1, and as s_i gets closer to 0.

(c) Adoption by all banks in some countries and by no banks in other countries. This can occur for any $\delta < 1$, if $\frac{c}{b}$ and the market structure meet certain criteria. As δ goes to 0, any market structure where the largest player in some countries is larger than $\frac{c}{b}$ while in other countries it is smaller than $\frac{c}{b}$, may adopt this equilibrium.

(d) *Adoption by some but not all banks in a country is not a Nash equilibrium.*

Proof. Part (a). Adoption by everybody gives all banks per customer profit of $b - c$ and since they can never do better than that, this is Nash equilibrium.

Part (b). Using the notation introduced earlier, let $q_i = \delta(1 - s_i)$ denote the share of foreign transactions for country i. Since transaction patterns are random within a country, an individual bank j in country i will also have a proportion q_i of its transactions be international. Domestic transactions represent a share $1 - q_i$ of the bank's transactions, and a share r_{ij} of these domestic transactions are between its own customers. The bank's in-house transactions are thus a fraction $(1 - q_i)r_{ij}$ of its total transactions. If this bank is the only one in the whole world to adopt g, it can only use g for exactly these in-house transactions and its profits will be equal to:

$$\begin{aligned} & (1 - q_i)r_{ij}b - c \\ = \; & [1 - \delta(1 - s_i)]r_{ij}b - c. \end{aligned} \tag{A.3}$$

Since this expression is increasing in r_{ij}, it is the largest bank in each country that will gain the most from adopting g. Let r_{i1} denote the domestic share of this largest bank in country i. Thus adoption by no bank is a Nash equilibrium

$$\begin{aligned} [1-\delta(1-s_i)]r_{i1}b - c &< 0 \text{ for all } i \Leftrightarrow \\ r_{i1} &< \frac{c}{b}\frac{1}{[1-\delta(1-s_i)]}. \end{aligned} \tag{A.4}$$

Taking partial derivatives we get:

$$\begin{aligned} \frac{\partial r_{i1}}{\partial \delta} &= \frac{c}{b}\frac{1-s_i}{[1-\delta(1-s_i)]^2} > 0 \\ \frac{\partial r_{i1}}{\partial s_i} &= \frac{c}{b}\frac{-\delta}{[1-\delta(1-s_i)]^2} < 0. \end{aligned}$$

Therefore the upper boundary for r_{i1} (and the parameter space where adoption by none can occur) increases for higher δ and lower s_i.

Alternatively we can write expression (A.4) as follows:

$$s_i < \left(\frac{c}{b}\frac{1}{r_{i1}} - 1\right)\frac{1}{\delta} + 1 \quad \text{for all } i. \tag{A.5}$$

If there is only one bank in each country we have $r_{i1} = 1$, and (A.5) reduces to expression (3.2) in chapter 3.

Part (d). Once the largest bank in a country has adopted, all other banks will follow, since they will get profit of:

$$(1-q_i)(r_{ij} + r_{i1})b - c > (1-q_i)r_{ij}b - c \qquad j \neq 1.$$

Hence adoption by some but not all banks in a country cannot be Nash equilibrium.

Part (c). I now consider the situation where all banks in one or more countries adopt g, but nobody else does. This is an equilibrium if two conditions are met. First, the largest bank in at least one country has adopted g (after which the others followed), let this be country a:[1]

$$\text{Condition 1:} \qquad (1-q_a)r_{a1}b - c > 0\ .$$

The second condition requires that the largest bank of each other country does not gain from adoption. If all banks in country a have adopted g, the profit from adoption by the largest bank in another country has to be negative:

$$\text{Condition 2:} \qquad [(1-q_i)r_{i1} + q_i\frac{s_a}{1-s_i}]b - c < 0.$$

[1] The remainder of this proof considers the equilibrium where banks in only one country adopt g. The equations for the equilibria where more than one country, but not all, adopt g can be derived by taking s_a to be the joint share of all countries that have adopted g.

Combining the two conditions gives:

$$(1-q_i)r_{i1} + q_i\frac{s_a}{1-s_i} < \frac{c}{b} < (1-q_a)r_{a1} \text{ for all } i \neq a. \tag{A.6}$$

The difference between the right most and left most terms of the inequalities gives an indication the range of parameters b and c for which this third type of equilibrium can occur. This difference is equal to:

$$\begin{aligned} & (1-q_a)r_{a1} - (1-q_i)r_{i1} - q_i\frac{s_a}{1-s_i} \\ = \; & [1-\delta(1-s_a)]r_{a1} - [1-\delta(1-s_i)]r_{i1} - \delta s_a. \end{aligned} \tag{A.7}$$

For $\delta = 1$ the expression (A.7) reduces to:

$$s_a r_{a1} - s_i r_{i1} - s_a = -s_a(1-r_{1a}) - s_i r_{i1}.$$

Since this is negative for all market structures, the two conditions cannot be met for $\delta = 1$, in line with the earlier result that without autarky there are only two equilibria.

For $\delta < 1$ however, there are always market structures where the difference is positive and thus for at least some values of $\frac{c}{b}$ the third equilibrium can occur. The expression (A.7) is positive if:

$$r_{i1} < \frac{[1-\delta(1-s_a)]r_{a1} - \delta s_a}{1-\delta(1-s_i)} \text{ for all } i \neq a. \tag{A.8}$$

To see that there are always market structures that satisfy this criteria, note that if $r_{a1} = 1$ (country a has only one bank) the condition becomes:

$$r_{i1} < \frac{1-\delta}{1-\delta(1-s_i)} \text{ for all } i \neq a.$$

Since $0 < s_i < 1$, the term on the right side of the inequality ($\frac{1-\delta}{1-\delta(1-s_i)}$) is always smaller than 1. So if the other countries are sufficiently fragmented the condition is met. For $\delta \to 0$ the condition in expression (A.6) becomes:

$$r_{i1} < \frac{c}{b} < r_{a1} \text{ for all } i \neq a,$$

i.e. the more concentrated countries adopt (those where $r_{i1} > \frac{c}{b}$) the rest doesn't. ■

Appendix B
Proof of propositions in chapter 4

Deriving formulas in Table 4.1

Proof. For all four scenarios (outcomes of stage 1) the share of both firms is defined by the 'marginal consumer' for whom the two hedonic prices (corrected for network benefits) and transportation costs t, (where $t = 1$) are equal:[1]

$$s_i = \frac{\hat{p}_j - \hat{p}_i + 1}{2}. \tag{B.1}$$

Here $\hat{p}_j = p_j - s_g b$, where s_g is the overall share of g. We now look at each scenario in turn; because scenario 2 is the most complicated, we save it for last.

Scenario 1: neither party adopts. In this scenario $s_g = 0$ and $b = c = 0$, leading to the standard (symmetrical) Hotelling result:

$$\begin{aligned} p_i^* &= 1 \\ s_i^* &= \frac{1}{2} \\ \pi_i^* &= \frac{1}{2}. \end{aligned}$$

Scenario 3: both firms introduce incompatible versions of g. We get $s_g{=}s_i$ and we can rewrite (B.1) as:

$$\begin{aligned} s_i &= \frac{p_j - p_i - b(s_j - s_i) + 1}{2} \Leftrightarrow \text{ (because } s_j = 1 - s_i) \\ s_i &= \frac{p_j - p_i - b(1 - 2s_i) + 1}{2} \Leftrightarrow \\ s_i &= \frac{p_j - p_i + 1 - b}{2(1 - b)}. \end{aligned} \tag{B.2}$$

[1] This is obtained by substituting $t = 1$ in the original Hotelling formula $s_i = (p_j - p_i + t)/2t$

The price response function for firm i given p_j is derived by maximizing firm i profit: $\pi_i = (p_i - c)s_i$, with s_i given by B.2. The first order condition is:

$$\frac{\partial \pi_i}{\partial p_i} = s_i + (p_i - c)s_i' = 0 \Leftrightarrow$$
$$\frac{(p_j - p_i + 1 - b)}{2(1-b)} - \frac{(p_i - c)}{2(1-b)} = 0 \Leftrightarrow$$
$$p_i = \frac{p_j + c + 1 - b}{2},$$

and because of symmetry we get the same expression for p_j. Solving these two equations for prices gives the (symmetrical) equilibrium prices, shares and profits for $i = 1, 2$:

$$\begin{aligned} p_i^* &= 1 - b + c \\ s_i^* &= \frac{1}{2} \\ \pi_i^* &= \frac{1}{2} - \frac{b}{2}. \end{aligned}$$

Scenario 4: both firms introduce compatible versions. Now $s_j = 1$, and

$$\begin{aligned} s_i &= \frac{\hat{p}_j - \hat{p}_i + 1}{2} = \frac{(p_j - b) - (p_i - b) + 1}{2} = \frac{p_j - p_i + 1}{2} \\ \pi_i &= (p_i - c)s_i. \end{aligned}$$

This leads us back to the equations of scenario 1, except that marginal costs for both firms have been raised by c. I then get the following equilibrium values for $i = 1, 2$:

$$\begin{aligned} p_i^* &= 1 + c \\ s_i^* &= \frac{1}{2} \\ \pi_i^* &= \frac{1}{2}. \end{aligned}$$

Scenario 2: only one firm introduces the new technology. The solution is asymmetrical and the expressions are more complicated. Without loss of generality, I assume firm 1 unilaterally adopts g. We get $\hat{p}_1 = p_1 - bs_1$ and $\hat{p}_2 = p_2$. Substituting this into (B.1), we get:

$$\begin{aligned} s_1 &= \frac{p_2 - p_1 + bs_1 + 1}{2} \Leftrightarrow \\ s_1 &= \frac{p_2 - p_1 + 1}{2 - b} \text{ and (since } s_2 = 1 - s_1) \Leftrightarrow \\ s_2 &= \frac{p_2 - p_1 - b + 1}{2 - b}. \end{aligned}$$

The price response for firm 1 given p_2 is derived by maximizing firm 1 profit: $\pi_1 = (p_1 - c)s_1$. The first order condition is:

$$\begin{aligned} \frac{\partial \pi_1}{\partial p1} &= s_1 + (p_1 - c)s_1^{'} = 0 \Leftrightarrow \\ p_1 &= \frac{p_2 + c + 1}{2}. \end{aligned} \tag{B.3}$$

The price response for firm 2 given p_1 is derived by maximizing $\pi_2 = p_2 s_2$. The first order condition is:

$$\begin{aligned} \frac{\partial \pi_2}{\partial p_2} &= s_2 + p_2 s_2^{'} \Leftrightarrow \\ p_2 &= \frac{p_1 - b + 1}{2}. \end{aligned} \tag{B.4}$$

The solution to equations (B.3) and (B.4) is given by:

$$\begin{aligned} p_1^* &= 1 - \frac{b - 2c}{3} \\ p_2^* &= 1 - \frac{2b - c}{3} \\ s_1^* &= \frac{1}{2} + \frac{b - 2c}{6(2 - b)} \\ s_2^* &= \frac{1}{2} - \frac{b - 2c}{6(2 - b)} \\ \pi_1^* &= \frac{(1 - \frac{b+c}{3})^2}{2 - b} \\ \pi_2^* &= \frac{(1 - \frac{2b-c}{3})^2}{2 - b}. \end{aligned}$$

■

Proposition 4.1 *"For small to moderate network effects, 2 players will share the market in equilibrium": iff* b< 1 *then stage 2 of the game will always yield an internal solution, where both players have positive market share.*

Proof. We show that $b < 1$ corresponds to the condition for stable internal solution formulated by DePalma and Leruth (1993). They formulate a demand function D:

$$Q_i = D\left[\frac{p_j - p_i}{\mu}, \frac{E_i - Ej}{\mu}\right], \tag{B.5}$$

where μ is a measure of differentiation, and E_i is the size of the externality for network i, where $E_i = R_I Q_i - R_E(1 - Q)$; R_I is the value of a subscriber

inside the network and R_E is the value of a consumer *outside* the network; E_i is therefore the benefit to network i of one extra subscriber. Demand is thus a function of relative prices and relative network benefits of both networks.

DePalma and Leruth show that (B.5) has a unique solution iff D satisfies

$$D_2[.] < \frac{\mu}{2(R_I - R_E)}, \tag{B.6}$$

where $D_2[.]$ is the derivative of the demand function with respect to its second argument. The idea behind this is as follows. If a customer switches from network 2 to 1, the externality (or network value) of network 1 goes up by $R_I - R_E$ (difference in value between a subscriber inside and outside the network). At the same time the externality of network 2 decreases by that same amount (the customer leaves network 2). If this change in externality changes the demand (=size of network 1) by more than one customer, the process feeds on itself: because of this additional customer, the difference in externality increases enough to raise demand by again 1 customer or more etc.

We now translate the demand function and condition to the parameters of our model. μ corresponds to our transportation costs t (which we have normalized to 1). Since we have normalized the number of consumers to one, their Q_i (number of users of network i) corresponds to our s_i, Their R_I (value of a subscriber in the network) is our parameter b, while we have set $R_E = 0$. Thus their parameter $E_i = R_I Q_I - R_E(1 - Q_I)$ corresponds to bs_i of our model. We need to take the derivative of our demand function with respect to their second argument:

$$\frac{E_i - E_j}{\mu} = b\frac{s_j - s_i}{t}.$$

In our model the demand function is given by (B.2), which we can rewrite as

$$s_i = \frac{p_j - p_i + b(s_i - s_j) + 1}{2}.$$

so that the derivative with respect to the second argument in (B.5) is $\frac{1}{2}$. Thus condition (B.6) becomes $\frac{1}{2} < t/2b$ or $b < 1$, since I normalized $t = 1$. ■

Proposition 4.1.1 *"For small or moderate network effects, competing on standards doesn't pay":*

(a) *If 2 players maintain incompatible versions of* g, *the profit for both of them is lower than under any other outcome of stage 1. As a result maintaining incompatible standards is not an equilibrium outcome.*

(b) *If one player unilaterally adopts* g, *profits of both firms go down.*

Proof. Part (a). Equilibrium profits for both firms under incompatibility are $\frac{1}{2} - \frac{b}{2}$, lower than the profit under either compatibility or non-adoption by both. Therefore it needs to be shown that profits under unilateral adoption are higher than $\frac{1}{2} - \frac{b}{2}$. Without loss of generality, assume firm 1 adopts, while firm 2 doesn't. Let π_1 and π_2 be the corresponding profits. I show that $\pi_1 > \frac{1}{2} - \frac{b}{2}$ ($\pi_2 > \frac{1}{2} - \frac{b}{2}$ follows using exactly the same approach).

$$\pi_1 = \frac{(1 - \frac{b+c}{3})^2}{2-b} > \frac{1}{2} - \frac{b}{2} \Leftrightarrow$$

$$c < 3 - b - 3\sqrt{\frac{(1-b)(2-b)}{2}} \text{ or } c > 3 - b + 3\sqrt{\frac{(1-b)(2-b)}{2}}. \qquad \text{(B.7)}$$

It can be verified that, as long as $0 < b < 1$:

$$b < 3 - b - 3\sqrt{\frac{(1-b)(2-b)}{2}}.$$

Thus the first inequality in (B.7) is satisfied, since $c < b$.

Part (b). Using the same definition of π_1 and π_2 as in part (a), I now need to show that both are lower than $\frac{1}{2}$, the profits under either non-adoption or adoption of compatible versions.

I show that $\pi_1 < \frac{1}{2}$ ($\pi_2 < \frac{1}{2}$ follows using exactly the same approach):

$$\pi_1 = \frac{(1 - \frac{b+c}{3})^2}{2-b} < \frac{1}{2} \Leftrightarrow$$

$$3 - b - 3\sqrt{1 - \frac{b}{2}} < c < 3 - b + 3\sqrt{1 - \frac{b}{2}}. \qquad \text{(B.8)}$$

It can be verified that, as long as $0 < b < 1.5$:

$$3 - b - 3\sqrt{1 - \frac{b}{2}} < 0 \text{ and } 3 < 3 - b + 3\sqrt{1 - \frac{b}{2}}.$$

So the conditions (B.8) are always satisfied for $0 < c < b < 1$. ■

Proposition 4.3 *For $\delta < 1$ and $b < 1$, the outcomes of the game are as follows:*

(a) *If $\frac{4(1-\delta)}{(3-2\delta)^2} > b > \frac{1}{2-\delta}$ then duopolists will prefer compatibility.*

(b) *If $b < \frac{4(1-\delta)}{(3-2\delta)^2}$ then firms prefer incompatibility over compatibility, even in the absence of migration costs.*

(c) *If $b > \frac{1}{2-\delta}$ then the DePalma-Leruth condition for coexistence of incompatible networks is no longer met, and the system tips to either standard.*

Proof. The proof consists of three parts. I first derive market shares as a function of p_1 and p_2. I then calculate equilibrium prices and profits as a function of b, c and δ. Finally, I derive the actual result of the proposition.

1. Market share functions. If $\delta < 1$, the market share functions are no longer continuous, if firms have incompatible standards. To see that, take the perspective of firm 2, as it tries to expand market share beyond $s_2 = \frac{1}{2}$. Attracting the nearest customer in the other 'country' requires a discontinuous lowering of price, since the network benefits for that customer are limited: he interacts mostly with customers of country 1, who use the incompatible network of firm 1. The network benefits of the 'last' customer in country 2 are equal to:

$$\left(1 - \frac{\delta}{2}\right) b. \tag{B.9}$$

since he performs a fraction $\frac{\delta}{2}$ with 'foreign' customers and thus a fraction $\left(1 - \frac{\delta}{2}\right)$ with domestic customers, and only domestic customers use his network. The network benefits of the first 'foreign' customer of firm 2 (located in country 1) are equal to:

$$\frac{\delta}{2} b. \tag{B.10}$$

since this customer can only use the network of firm 2 for his foreign transactions. Thus, to attract the first foreign customer, the price needs to drop by the difference between (B.9) and (B.10) or:

$$\left(1 - \frac{\delta}{2}\right) b - \frac{\delta}{2} b = (1 - \delta) b.$$

Hence all prices that satisfy $|p_1 - p_2| \leq (1 - \delta)b$ will lead to $s_1 = s_2 = \frac{1}{2}$ (note that if $\delta = 1$, the required drop is 0, and the function is continuous). To get market shares if $|p_1 - p_2| > (1 - \delta)b$, first assume $p_2 < p_1 - (1 - \delta)b$. In that case the marginal customer, who is indifferent between firm 1 and 2, will reside

in country 1. By definition the address of this marginal customer is s_1. Of all his transactions, a share $\left(1-\frac{\delta}{2}\right)$ is with customers in country 1. Of these, a fraction $2s_1$ are using the network of firm 1. Hence a share of his transactions equal to $\left(1-\frac{\delta}{2}\right)2s_1 = (2-\delta)s_1$ is with customers on network 1 and $1-(2-\delta)s_1$ is with customers on network 2. Now for this marginal customer the benefits of both networks need to be the same:

$$p_1 + ts_1 - b(2-\delta)s_1 = p_2 + t(1-s_1) - b\left[1-(2-\delta)\,s_1\right].$$

Normalizing $t = 1$ and solving for s_1:

$$\begin{aligned} 2s_1 - 2b\,(2-\delta)\,s_1 &= p_2 - p_1 + 1 - b \Leftrightarrow \\ s_1 &= \frac{p_2 - p_1 + 1 - b}{2 - 2b(2-\delta)}, \end{aligned} \tag{B.11}$$

and

$$s_2 = 1 - s_1 = \frac{p_1 - p_2 + 1 - b(3-2\delta)}{2 - 2b(2-\delta)}.$$

Using a similar approach for the case where the marginal customer is in country 2, we get the following share function for s_1 (and $s_2 = 1 - s_1$):

$$\begin{aligned} &\text{if } |p_1 - p_2| \le (1-\delta)b\text{:} && s_1 = \frac{1}{2} \\ &\text{if } p_1 - p_2 > (1-\delta)b\text{:} && s_1 = \frac{p_2 - p_1 + 1 - b}{2 - 2b(2-\delta)} \\ &\text{if } p_2 - p_1 > (1-\delta)b\text{:} && s_1 = \frac{p_2 - p_1 + 1 - b(3-2\delta)}{2 - 2b(2-\delta)}. \end{aligned} \tag{B.12}$$

2. Equilibrium prices and profits. We now get a situation where a firm can only expand its market share above $\frac{1}{2}$ if it substantially drops its price, 'undercutting' its rival. To analyze this type of situation I use the concept of Undercut Proof Equilibrium (UPE), described in Shy (2002). It defines equilibrium as a situation where neither firm can profitably undercut its rival's price.[2] Let π_i^*, p_i^* and s_i^* denote UPE profits, prices and shares. Because of the symmetry $s_1^* = s_2^* = \frac{1}{2}$. Suppose firm 2 tries to undercut its rival, who is charging p_1^*, by offering a price p_2. We can now maximize firm 2 profits: $\pi_2 = (p_2 - c)s_2$. The

[2] Technically this is a broader definition than Nash-equilibrium. Nash-equilibrium requires that neither player can unilaterally improve his results by any price change. UPE requires that neither player can unilaterally improve his results by *lowering* price. Generally, in UPE, both players can unilaterally improve profits by *increasing* prices.

first order condition is:

$$\begin{aligned} \frac{\partial \pi_2}{\partial p_2} &= s_2 + (p_2 - c)s_2' = 0 \Leftrightarrow \\ p_1^* - p_2 + 1 - b(3 - 2\delta) &= p_2 - c \Leftrightarrow \\ p_2 &= \frac{p_1^* + 1 - b(3 - 2\delta) + c}{2}. \end{aligned}$$

If firm 2 indeed selects this price p_2 we get:

$$\begin{aligned} s_2 &= \frac{p_1^* - p_2 + 1 - b(3 - 2\delta)}{2 - 2b(2 - \delta)} = \frac{\frac{p_1^* + 1 - b(3-2\delta) - c}{2}}{2 - 2b(2 - \delta)} \\ \pi_2 &= (p_2 - c)s_2 = \frac{\left[\frac{p_1^* + 1 - b(3-2\delta) - c}{2}\right]^2}{2 - 2b(2 - \delta)}. \end{aligned}$$

If firm 2 were to charge $p_2^* = p_1^*$ (instead of undercutting firm 1), its profit would be $(p_1^* - c)\frac{1}{2}$. So a price p_1^* is undercut proof if firm 2 cannot get a higher profit by deviating form $p_2 = p_1^*$:

$$\begin{aligned} \frac{(p_1^* - c)}{2} &\geq \pi_2 \\ \frac{(p_1^* - c)}{2} &\geq \frac{\left[\frac{p_1^* + 1 - b(3-2\delta) - c}{2}\right]^2}{2 - 2b(2 - \delta)} \end{aligned}$$

We can directly substitute $\pi_1^* = (p_1^* - c)\frac{1}{2}$ to obtain:

$$\begin{aligned} [2 - 2b(2 - \delta)]\pi_1^* &\geq \left[\pi_1^* + \frac{1 - b(3 - 2\delta)}{2}\right]^2 \Leftrightarrow \\ 0 &\geq (\pi_1^*)^2 - (1 - b)\pi_1^* + \left[\frac{1 - b(3 - 2\delta)}{2}\right]^2. \end{aligned}$$

π_1^* is the highest value that satisfies this quadratic inequality. Hence:

$$\begin{aligned} \pi_1^* &= \frac{1 - b + \sqrt{D}}{2} \text{ with} \\ D &= (1 - b)^2 - [1 - b(3 - 2\delta)]^2. \end{aligned}$$

Because of symmetry, $\pi_1^* = \pi_2^*$. Note that if $\delta = 1$ (the case analyzed in section 3) we get $D = 0$ and $\pi_i^* = \frac{1}{2} - \frac{b}{2}$, which is exactly the result obtained in section 3 (table 1). Hence the concept of UPE converges to the standard Hotelling Nash-equilibrium as $\delta \to 1$ (as it should).

3. Deriving the actual result of the proposition. We now look for values of b and δ where π_i^* is larger than the profit under compatibility, which is $\frac{1}{2}$:

$$\begin{aligned}
\frac{1}{2} &< \frac{1-b+\sqrt{D}}{2} \Leftrightarrow \\
b^2 &< D \Leftrightarrow \\
0 &< -2b+6b-4b\delta-b^2(3-2\delta)^2 \Leftrightarrow \\
0 &< 4-4\delta-b(3-2\delta)^2 \Leftrightarrow \\
b &< \frac{4(1-\delta)}{(3-2\delta)^2}.
\end{aligned}$$

This last relationship is plotted in figure 4.1. The requirement $b < \frac{1}{2-\delta}$ follows from the fact that an internal solution only exists if the denominator in (B.11) is positive. ■

Proposition 4.4 *If* $\alpha \geq 0$, *and* b< 1 *(network effects are not too large) there are two equilibria:*

(a) *Adoption of compatible versions. This an equilibrium for all values of* α *and* $\frac{c}{b}$.

(b) *Non-adoption by both firms is an equilibrium if* $\frac{c}{b}$ *is high and/or* α *is low, according to the curve in figure 4.3.*

(c) *Neither incompatible versions nor partial adoption (one firm adopts, the other doesn't) are an equilibrium for any value of the parameters* 0 <c<b *and* $\varepsilon \geq 0$.

Proof. The area below the curve in figure 4.3 corresponds to parameter values where $\pi_i^{nobody-adopts} < \pi_i^{one-adopts} < \pi_i^{compatible}$ for $i = 1, 2$. To find that area I first analytically derive equilibrium prices and profits for the three symmetrical outcomes: (nobody adopts, both adopt compatible versions and both adopt incompatible versions), with $\varepsilon \geq 0$. I then describe how I numerically derived profits for the fourth (asymmetrical) case: unilateral adoption by only 1 firm.

1. Both players adopt compatible versions. If both players adopt compatible versions of g, we get:

$$\pi_i = [p_i D(p_i) - c]s_i. \tag{B.13}$$

Which is a third order polynomial in p_i. The first order condition is:

$$\begin{aligned} \frac{\partial \pi_i}{\partial p_i} &= [D(p_i) + p_i D'(p_i)]s_i + [p_i D(p_i) - c]s_i' \\ & [1 + \alpha(1 - p_i + b) - \alpha p_i]s_i - \frac{p_i}{2}[1 + \alpha(1 - p_i + b)] + \frac{c}{2} = 0. \end{aligned}$$

Here $D'(p_i) = -\alpha$ and $s_i' = -\frac{1}{2}$ denote first derivatives with respect to p_i. Both systems are compatible so we have $s_g = 1$. Because the equilibrium is symmetrical we can substitute $p^* = p_i = p_j$ and $s^* = s_1 = s_2 = \frac{1}{2}$:

$$[1 + \alpha(1 - p^* + b) - \alpha p^*]\frac{1}{2} - \frac{p^*}{2}[1 + \alpha(1 - p^* + b)] + \frac{c}{2} = 0. \tag{B.14}$$

This is a quadratic expression in p^* that can be solved analytically (but the resulting expression is too long to reproduce here). By substituting p^*in (B.13) we get equilibrium profits.

2. Neither player adopts: Equilibrium prices and profits follow by simply taking $b = c = 0$ in expression (B.14).

3. Both players adopt incompatible versions. Because $D(p_i) = 1+\alpha(1-p+s_i b)$ and $s_i = \frac{p_j - p_i + 1 - b}{2(1-b)}$, we now have $D'(p_i) = -\alpha - \frac{b}{2(1-b)}$ and the first order condition becomes:

$$\frac{\partial \pi_i}{\partial p_i} = [D(p_i) + p_i D'(p_i)]s_i + [p_i D(p_i) - c]s_i' = 0$$

$$\Leftrightarrow \quad [1 + \alpha(1 - p_i + s_i b) - (\alpha + \frac{b}{2(1-b)})p_i]s_i - \frac{p_i[1 + \alpha(1 - p_i + s_i b)] - c}{2(1-b)}b = 0.$$

Because the equilibrium is symmetrical we can substitute $p^* = p_i = p_j$ and $s^* = s_1 = s_2 = \frac{1}{2}$:

$$[1 + \alpha(1 - p^* + \frac{b}{2}) - (\alpha + \frac{b}{2(1-b)})p^*](1-b) - p^*[1 + \alpha(1 - p^* + \frac{b}{2})] + \frac{c}{2} = 0.$$

Which is again a solvable quadratic expression in p^*.

4. One player adopts, while the other doesn't. This case can in theory be approached analytically, but the expressions become very complex. Hence they were solved numerically using a pascal program which is available upon request. The results in figure 4.3 were numerically generated, by comparing the outcomes. ■

Proposition 4.5 *Given natural shares $\hat{s}_i$, equilibrium is given by:*

$$\begin{aligned} p_i^* &= \frac{1}{(1-k)(2-\hat{s}_i)} \text{ where } k \equiv \sum_{j=1}^{n} \frac{\hat{s}_j}{2-\hat{s}_j} \\ s_i^* &= \hat{s}_i \left(\frac{1}{1-k}\right)\left(\frac{1-\hat{s}_i}{2-\hat{s}_i}\right) \\ \pi_i^* &= p_i^* s_i^* = \hat{s}_i \left(\frac{1}{1-k}\right)^2 \frac{1-\hat{s}_i}{(2-\hat{s}_i)^2}. \end{aligned} \tag{B.15}$$

Proof. The actual share of firm i is determined endogenously by prices and by the natural shares ($\hat{s}_i$):

$$s_i = (\hat{s}_i)^2 + \sum_{j \neq i} \hat{s}_i \hat{s}_j (p_j - p_i + 1).$$

The first term represents the market where firm i competes against itself, the second term represents the markets where it competes against each of its competitors. We can rewrite this equation as:

$$s_i = \hat{s}_i(\overline{p} - p_i + 1), \text{where } \overline{p} \equiv \sum_{i=1}^{n} \hat{s}_i p_i. \tag{B.16}$$

Profits are given by $\pi_i = p_i s_i = p_i \hat{s}_i (\overline{p} - p_i + 1)$. The first order condition is

$$\begin{aligned} \frac{\partial \pi_i}{\partial p_i} &= \hat{s}_i(\overline{p} - p_i + 1) + p_i \hat{s}_i(\hat{s}_i - 1) = 0 \Leftrightarrow \\ p_i &= \frac{\overline{p} + 1}{(2 - \hat{s}_i)}. \end{aligned}$$

Let p_i^* denote the equilibrium prices for firm i and $\overline{p}^*$ the average equilibrium price across all firms. Now we can rewrite

$$\begin{aligned} \overline{p}^* &= \sum_{i=1}^{n} \hat{s}_i p_i^* = \sum_{i=1}^{n} \hat{s}_i \left(\frac{\overline{p}^* + 1}{(2 - \hat{s}_i)}\right) \Leftrightarrow \\ \frac{\overline{p}^*}{\overline{p}^* + 1} &= \sum_{i=1}^{n} \frac{\hat{s}_i}{(2 - \hat{s}_i)} \equiv k. \end{aligned}$$

For ease of notation I denote this last term k and thus:

$$\begin{aligned} \overline{p}^* &= \frac{k}{1-k} \text{ and} \\ p_i^* &= \frac{\overline{p}^* + 1}{(2 - \hat{s}_i)} = \frac{1}{(1-k)(2-\hat{s}_i)}. \end{aligned}$$

Entering these prices in (B.16) gives equilibrium market shares and profits:

$$\begin{aligned} s_i^* &= \hat{s}_i(\overline{p}^* - p_i^* + 1) = \hat{s}_i \left(\frac{1}{1-k}\right)\left(\frac{1-\hat{s}_i}{2-\hat{s}_i}\right) \\ \pi_i^* &= p_i s_i = \hat{s}_i \left(\frac{1}{1-k}\right)^2 \frac{1-\hat{s}_i}{(2-\hat{s}_i)^2}. \end{aligned}$$

■

Proposition 4.6 *Compared to the model for unsponsored standards, the availability of proprietary standards increases the range of parameter values where suboptimal equilibria can occur. This holds for fixed demand in the following industry structures: (i) any duopoly, (ii) any Gorilla vs. competitive fringe, (iii) any number of equal sized firms.*

Proof. Profits for coalitions of players were calculated as follows. Let there be two coalitions: coalition 1 with m members, and coalition 2 with $n-m$ members. I use the following notation:

$$
\begin{aligned}
c_i &: \text{denotes each of the two coalitions } i = 1,2 \;\; c_1 = \{1,..,m\},\, c_2 = \{m+1,..,n\} \\
\overline{p} &\equiv \sum_{i=1}^{n} \hat{s}_j p_j : \text{average price of all firms} \\
\overline{p}_{c_i} &\equiv \sum_{j \epsilon c_i} \hat{s}_j p_j : \text{part of the average price belonging to members of } c_i \\
\hat{p}_i &\equiv p_i - b s_g : \text{hedonic price, } s_g \text{ is the size of the network} \\
\widehat{\overline{p}} &\equiv \sum_{i=1}^{n} \hat{s}_i \hat{p}_i : \text{the average hedonic price} \\
\hat{s}_{c_i} &\equiv \sum_{j \epsilon c_i} \hat{s}_j : \text{natural market share of coalition } i \\
s_{c_i} &\equiv \sum_{j \epsilon c_i} s_j : \text{actual market share of coalition } i
\end{aligned}
$$

The hedonic prices can then be inserted in (B.16) to derive market shares given all prices. However, since these hedonic prices in turn depend on s_g, which depends on the (endogenous) market shares s_i (as opposed to the exogenous natural shares $\hat{s}_i$), some additional manipulations are needed to derive closed form equations for s_i. This is done below, for two cases: (1) only one coalition adopts the technology, (2) both coalitions adopt incompatible versions.[3]

1. Just coalition 1 adopts the technology. Corrected prices are then:

$$
\begin{aligned}
\hat{p}_i &= p_i - b s_{c1} \text{ for } i \epsilon c_1 \\
\hat{p}_i &= p_i \text{ for } i \epsilon c_2 \\
\widehat{\overline{p}} &= \sum_{i=1}^{n} \hat{s}_i \hat{p}_i = \sum_{i \epsilon c_1} \hat{s}_i (p_i - b s_{c1}) + \sum_{j \epsilon c_2} \hat{s}_j p_j = \overline{p} - \hat{s}_{c_1} b s_{c_1}
\end{aligned}
$$

By substituting the expressions for $\hat{p}_i$ and $\widehat{\overline{p}}$ in (B.16) we get:

[3] If both coalitions adopt compatible versions, market shares are again given by the basic equation (B.16).

$$
\begin{aligned}
s_i &= \hat{s}_i(\overline{p} - \hat{s}_{c_1} b s_{c_1} - p_i + b s_{c_1} + 1) = \hat{s}_i(\overline{p} - p_i + \hat{s}_{c_2} b s_{c_1} + 1) \Leftrightarrow \\
s_{c_1} &= \sum_{i \epsilon c_1} s_i = \hat{s}_{c_1}(\overline{p} + \hat{s}_{c_2} b s_{c_1} + 1) - \overline{p}_{c_1} \Leftrightarrow \\
s_{c_1} &= \frac{\hat{s}_{c_1}(\overline{p} + 1) - \overline{p}_{c_1}}{1 - \hat{s}_{c_1}\hat{s}_{c_2} b}. \qquad \text{(B.17)}
\end{aligned}
$$

By substituting this value of s_{c_1} into (B.17) we get the actual shares for the members of coalition 1 as a function of prices. The shares for the members of coalition 2 (who do not adopt g) are given by:

$$s_j = \hat{s}_j(\overline{p} - \hat{s}_{c_1} b s_{c_1} - p_j + 1).$$

with s_{c_1} as given by (B.17). Using these share functions, I then iteratively determined equilibrium prices and profits; one member of each coalition chooses the price that maximizes profits, given the prices of all players. In the next iteration all other members follow this price, after which the first member again optimizes price, etc., until convergence.

2. Both coalitions adopt. Corrected prices are then:

$$
\begin{aligned}
\hat{p}_i &= p_i - b s_{c_i} \\
\widehat{\overline{p}} &= \sum_{i \epsilon c_1} \hat{s}_i \hat{p}_i + \sum_{j \epsilon c_2} \hat{s}_j \hat{p}_j = \overline{p} - \hat{s}_{c_i} b s_{c_i} - \hat{s}_{c_2} b s_{c_2}.
\end{aligned}
$$

Again substituting $\hat{p}_i$ and $\widehat{\overline{p}}$ in (B.16) we get:

$$s_i = \hat{s}_i(\overline{p} - \hat{s}_{c_1} b s_{c_1} - \hat{s}_{c_2} b s_{c_2} - p_i + b s_{c_i} + 1).$$

For $i \in c_1$ this becomes:

$$
\begin{aligned}
s_i &= \hat{s}_i(\overline{p} + \hat{s}_{c_2} b s_{c_1} - \hat{s}_{c_2} b s_{c_2} - p_i + 1) \\
&= \hat{s}_i(\overline{p} - p_i + \hat{s}_{c_2} b(1 - 2 s_{c_1}) + 1) \text{ for } i \epsilon c_1. \qquad \text{(B.19)}
\end{aligned}
$$

Summing over all i in coalition 1, we get:

$$s_{c_1} = \frac{\hat{s}_{c_1}(\overline{p} + \hat{s}_{c_2} b - 1) - \overline{p}_{c_1}}{1 - 2\hat{s}_{c_1}\hat{s}_{c_2} b}. \qquad \text{(B.20)}$$

By substituting this value of s_{c_1} into (B.19) we get the actual shares for the members of coalition 1. The shares for the members of coalition 2 follow by symmetry. ■

Proposition 4.7 *For moderate network effects, and all $\alpha \geq 0$, the adoption of compatible versions of* g *by both firms is the socially optimal outcome.*

Proof. Social welfare is the sum of firm welfare and consumer welfare: $W_S = W_F + W_C$. $W_F = \sum \pi_i^*$ can be calculated with the equations in table 4.1. To calculate W_C let $S(p_i)$ denote the surplus for an individual consumer given a price p_i. It is equal to:[4]

$$\begin{aligned} S(p_i) &= \int_{p_i}^{1+\frac{1}{\alpha}+s_g b} D(x)dx = \int_{p_i}^{1+\frac{1}{\alpha}+s_g b} (1+\alpha(1-x+s_g b)dx \\ &= \left[(1+\alpha+\alpha s_g b)x - \frac{\alpha}{2}x^2\right]_{p_i}^{\infty} \\ &= (1+\alpha+\alpha s_g b)(1+\frac{1}{\alpha}+s_g b-p_i) - \frac{\alpha}{2}(1+\frac{1}{\alpha}+s_g b-p_i)^2. \end{aligned}$$

Now W_C can be calculated as follows:

$$W_C = \sum_{i=1,2} S(p_i)s_i.$$

Using the equations for W_F and W_C I performed a numerical grid search (using an excel spreadsheet program) for the following parameters:

- α from 0 to 10 in 0.1 increments. In addition I checked $\alpha = 1000$.
- b from 0 to 1 in 0.1 increments.
- $\frac{c}{b}$ from 0 to 1 in 0.1 increments.
- s_g follows the various outcomes of stage 1 of the game: (1) $s_g = 0$ if neither adopts; (2) $s_g = 1$ if both adopt compatible versions; and (3) $s_g = \frac{1}{2}$ if both adopts incompatible versions. For the asymmetrical outcome (only one firm adopts) s_g for the adopting firm is equal to its market share while $s_g = 0$ for the other firm. These shares were first calculated using the procedures outlined earlier, after which W_F and W_C were calculated.

In all of these cases did full adoption yield the largest social welfare. ■

[4] I take the integral from p_i to $1 + \frac{1}{\alpha} + s_g b$, because $D(p) = 0$ for $p = 1 + \frac{1}{\alpha} + s_g b$; for larger prices demand becomes negative.

Deriving table 4.7

For firm i the minimax strategy for the stage 2 pricing game is defined as:

$$\underset{a_i}{Max}\left(\underset{a_j}{Min}\right) o_{ij}.$$

Here a_i and a_j are possible actions for firms i and j, while o_{ij} is the payoff to firm i given these actions. The minimax solution is then the outcome if both players follow their minimax strategy. Below I derive the minimax outcomes for each of the four situations that may come out of stage 1 of the game.

1. Neither firm adopts. For any price p_1 set by firm 1, the worst that can happen to firm 1 is that firm 2 sets $p_2 = 0$. In that case:

$$\begin{aligned} s_1 &= \frac{0 - p_1 + 1}{2} \\ \pi_1 &= \frac{1}{2}p_1 - \frac{1}{2}(p_1)^2. \end{aligned}$$

To maximize this worst, outcome firm 1 needs to set $p_1 = \frac{1}{2}$. The minimax outcome is then $\pi_1 = \pi_2 = \frac{1}{2}$.

2. Only one firm adopts. Assume this is firm 1. Again the worst that can happen to firm 1 is $p_2 = 0$, in which case:

$$\begin{aligned} s_1 &= \frac{0 - p_1 + 1}{2 - b} \\ \pi_1 &= (p_1 - c)s_1 = \frac{(p_1 - c)(1 - p_1)}{2 - b}. \end{aligned}$$

This is maximized if

$$\begin{aligned} \frac{\partial \pi_1}{\partial p_1} &= \frac{(1 - p_1) - (p_1 - c)}{2 - b} = 0 \Leftrightarrow \\ p_1 &= \frac{1 + c}{2}. \end{aligned}$$

In a similar way, the worst that can happen to firm 2 is that firm 1 sets $p_1 = c$. In that case the best firm 2 can do is:

$$p_2 = \frac{c - b + 1}{2}.$$

This leads to

$$s_1 = \frac{p_2 - p_1 + 1}{2 - b} = \frac{1 - \frac{1}{2}b}{2 - b} = \frac{1}{2}.$$

So the minimax outcome is

$$\begin{aligned}\pi_1 &= (p_1 - c)s_1 = \frac{1}{4} - \frac{c}{4} \\ \pi_2 &= p_2 s_2 = \frac{1}{4} - \frac{b-c}{4}.\end{aligned}$$

3. Both firms adopt incompatible versions. The worst that can happen to firm 1 is $p_2 = c$. Then

$$\begin{aligned}s_1 &= \frac{c - p_1 - b + 1}{2(1-b)} \\ \pi_1 &= (p_1 - c)s_1.\end{aligned}$$

π_1 is maximized for

$$p_1 = \frac{1-b}{2} - c.$$

Due to symmetry we get $p_2 = p_1$ and $s_1 = \frac{1}{2}$. Minimax is then:

$$\pi_1 = \frac{1}{4} - \frac{b}{4}.$$

4. Both firms adopt compatible versions. As before profits are the same as under non-adoption, so $\pi_i = \frac{1}{4}$.

Appendix C

Proof of propositions in chapter 8

Conjecture 8.2 *The value of several variables in the model can be approximated by the following formulas:*

1. *For* m= 1..10, *the probability that an individual technology is profitable is equal to:*

$$P(\pi_k > 0) = \frac{1}{2} + \frac{1}{2}\left[\sum_{k=1}^{m}\binom{m}{k}p^k(1-p)^{m-k}\right]^2 - \frac{1}{2}p^{2m}.$$

2. *Given* m *and* n, *and* p= $\frac{1}{2}$, *the average number of accessible equilibrium points is approximately equal to:*

$$\begin{aligned}\overline{a}(n,m) &\approx 2^{q+1}\sqrt{\frac{2q-3}{2\pi q}} \text{ with}\\ q &= \frac{2}{3}nP(\pi_k > 0) \text{ where } P(\pi_k > 0) \text{ is defined as in (8.4).}\end{aligned}$$

3. *Given* m *and* n, *the probability of multiple equilibria can be approximated by:*

$$P[a(m,n) > 1] \approx 1 - F_{Weibull(\alpha,\beta)}(1),$$

where $F_{Weibull(\alpha,\beta)}$ *is the cumulative Weibull distribution with parameters:* $\alpha = \frac{\overline{a}(n,m)}{\Gamma(1+\frac{1}{\beta})}$ *and* $\beta = 2.5$.[1]

Proof. Part 1. In principle, there is a 50% chance of a technology having more benefits than costs, except for the fact that if the number of benefit and cost components are equal (and positive) the technology is profitable, because I assume that the benefit elements have a slightly higher value than the cost

[1] Here Γ denotes the Gamma function: $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dt$. Since the mean of a Weibull (α,β) function is equal to $\alpha\Gamma(1+\frac{1}{\beta})$, we get the best Weibull fit by taking $\alpha = \frac{m*n}{\Gamma(1+\frac{1}{\beta})}$.

elements. The probability of an equal and positive number of benefit and cost elements is:

$$\sum_{k=1}^{m}\left[\binom{m}{k}p^k(1-p)^k\right]^2.$$

Since the starting point of 50% includes half of this borderline, we need to add half of the borderline to 50% and then subtract half of the 'equal and zero cases' to get the probability we are looking for.

$$P(\pi_k > 0) = \frac{1}{2} + \frac{1}{2}\left[\sum_{k=1}^{m}\binom{m}{k}p^k(1-p)^{m-k}\right]^2 - \frac{1}{2}p^{2m}.$$

For $p = \frac{1}{2}$ this term works out to 62.5% for $m = 2$, rises to 64.1% for $m = 3$ then slowly declines with m to 58.8% for $m = 10$ and 52.8% for $m = 100$ and 50% for $m \to \infty$.[2]

Part 2. To derive the second part of the theorem, I start with the total number of combinations of technologies; n technologies will give rise to the following number of combinations:[3]

$$\sum_{k=1}^{n}\binom{n}{k} \approx \sqrt{2n-3}\binom{n}{\frac{1}{2}n} \approx 2^{n+1}\sqrt{\frac{2n-3}{2\pi n}} \text{ (for } n \geq 3). \qquad \text{(C.1)}$$

This is generally an overestimate of the number of accessible equilibrium points. Many combinations of technologies are not paths (because they include unprofitable adoptions), or lead to unprofitable points or to points where another technology can be profitably adopted from that point. To get an approximation of the number of accessible equilibrium points, I assume that the number of accessible equilibrium points is related to the number of points that are reached by combining technologies that are *profitable on a stand-alone basis.* Of course this cannot be a direct relationship: not all combinations of profitable technologies lead to accessible equilibrium points and some equilibrium points are reached by adopting technologies that are not profitable on a standalone basis. Using the first part of the theorem the average number of profitable

[2] If I assume that the benefit elements have a value smaller than 1.1 but still bigger than 1, the same reasoning and the resulting formula hold for values of $m > 10$.

[3] The first approximation uses the fact that the number of combinations is the largest if we pick half the total number of items. The second uses Stirling's formula ($n! \approx \left(\frac{n}{e}\right)^n\sqrt{2\pi n}$) to rewrite:

$$\binom{n}{\frac{1}{2}n} = \frac{n!}{\frac{n}{2}!\frac{n}{2}!} \approx \frac{\left(\frac{n}{e}\right)^n}{\left(\left(\frac{n}{2e}\right)^{\frac{n}{2}}\right)^2}\frac{\sqrt{2\pi n}}{\left(\sqrt{2\pi\frac{n}{2}}\right)^2} = \frac{2^{n+1}}{\sqrt{2\pi n}}$$

technologies is equal to $nP(\pi_k > 0)$. Thus, by using this number as input n for equation C.1 we get a first estimate of the number of equilibria. Comparing this predicted number of equilibria with the actual number obtained from the Monte Carlo simulation, I find that this method indeed overestimates the number of equilibria. A better prediction is obtained by using only $\frac{2}{3}$rds of all profitable technologies, i.e. $q \equiv \frac{2}{3}nP(\pi_k > 0)$, as input for (C.1). This method yields a surprisingly accurate prediction of the average number of accessible equilibrium points for n in the range 3-10 (higher values of n do not lend themselves to Monte Carlo simulation) and $p = \frac{1}{2}$.

Part 3. For the third part of the theorem we also need the distribution of the number of accessible equilibrium points. Based upon the Monte Carlo exercise this distribution is best described by a Weibull distribution with parameters:[4]

$$\alpha = \frac{\overline{a}(m,n)}{\Gamma(1+\frac{1}{\beta})}, \quad \beta = 2.5.$$

The fit of this approximation is given in figure C.1. ■

[4] As will be noted further down the value of the β parameter depends on the value of p, the probability that a single element of **B** or **C** is 1. For $p = 0.5$ we have $\beta = 2.5$ and for $p = 0.3$ I get $\beta = 1.8$.

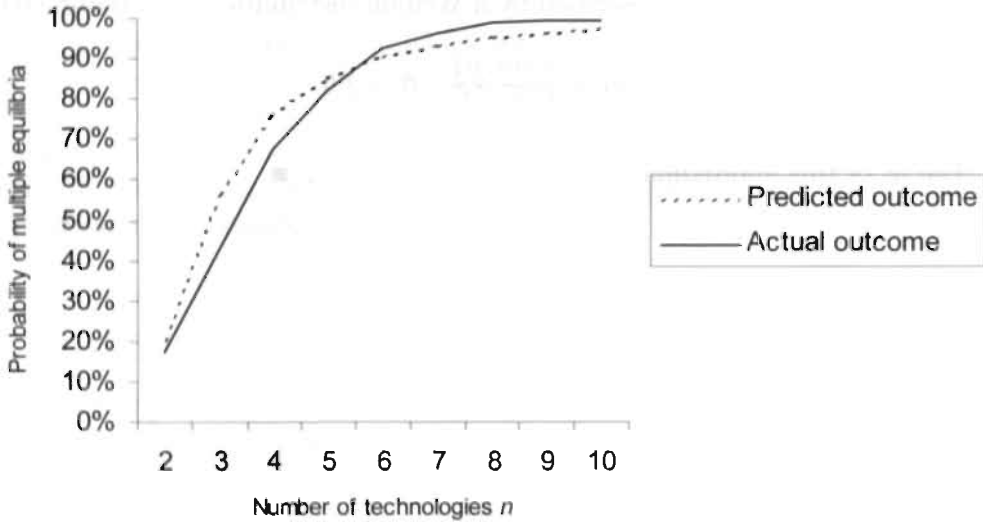

Figure C.1 Predicted versus actual probability of multiple accessible equilibrium points as a function of number of technologies n

Appendix D
Distribution of technical change sizes

Figure 8.14 shows the cost increases following the adoption of a technology. I have analyzed which statistical distribution best fits this pattern. The approach follows Law and Kelton (1991), chapter 6. I have fitted parameters for the following distributions: Normal, Lognormal, Poisson, Weibull and Gamma.[1] I then used the Chi-Square tests (described on p. 382-384 of that work) to test goodness of fit.

The results are given in table D.1. The values of the statistic for the Weibull distribution are significant at the 1% level, all others are not significant at any level. Since the test is not defined for the observations with value zero (there are no predicted observations for that value), the statistic leaves those out. Hence the curve for all values above zero is a Weibull with 99% confidence, but the zero-part is not. However, of all distributions, the Weibull distribution fits the pattern best for both small and large m. And as m increases, the importance of the zero-values declines. Figures D.1.and D.2 show the distributions of the increase in cost elements, as well as in profits for $m = 50$.

[1]The fitting was done using the best of MLE estimators given on p. 343-350 of Law and Kelton (1991) and visual fitting (best as measured by the χ^2-test). In addition I have visually fitted the binomial and powerlog distributions (for the tail only), neither of which gave a very good fit.

TABLE D.1 Results of fitting distribution to increases in costs

	m=10		m=50	
	$\mu(\sigma)$:$\alpha(\beta)$	χ^2	$\mu(\sigma)$:$\alpha(\beta)$	χ^2
Normal	1.40 (0.83)	2,613	5.83 (3.67)	8,944
Lognormal	0.33 (0.58)	2,213	1.71 (0.69)	3,363
Poisson	1.41	16,403	5.55 (3.28)	26,956
Gamma	2.78 (0.53)	552	3.28 (1.94)	2,325
Weibull	1.80 (1.60)	4.1	1.78 (6.91)	7.1

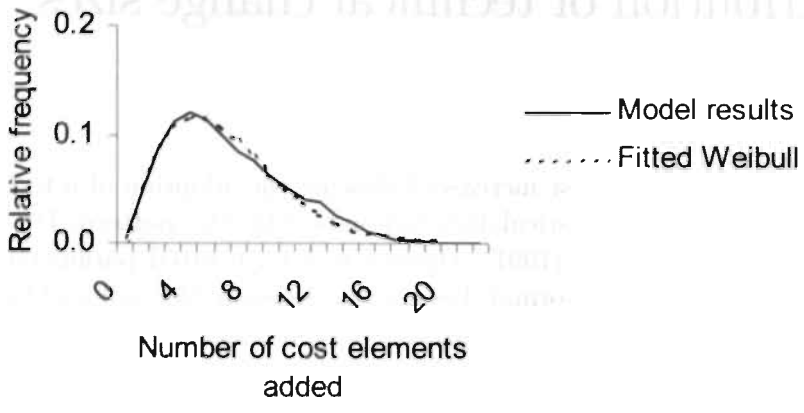

Figure D.1 Incremental costs per adoption for $m = 50$.

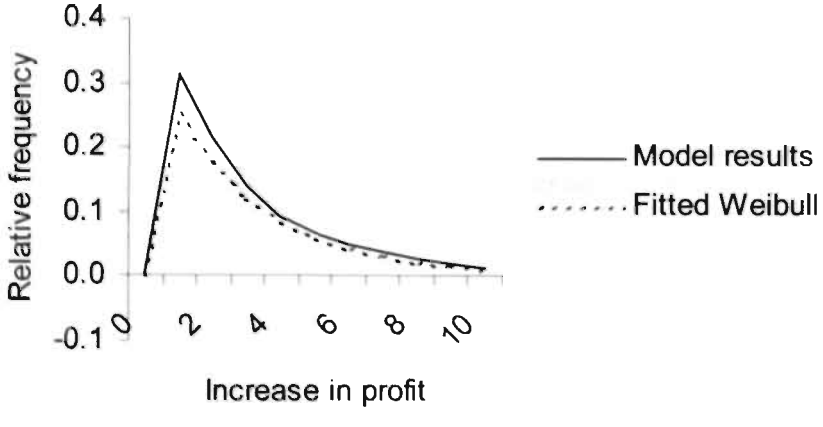

Figure D.2 Incremental profits per adoption for $m = 50$.

Summary in Dutch

Netwerk technologieën hebben zich de afgelopen jaren mogen verheugen in een warme belangstelling van economen. Het blijkt dat een groot aantal technologieën netwerk effect vertonen: voor elke gebruiker neemt de waarde van de technologie toe naarmate meer mensen gebruik maken van dezelfde standaard en hetzelfde netwerk. Tevens blijkt dat dergelijke technologieën een aantal bijzondere economische eigenschappen hebben. Zo vormen ze vaak een natuurlijk monopolie, waarbij er ruimte is voor slechts één standaard. Indien bedrijven standaarden of netwerken gesloten kunnen houden, beïnvloeden netwerk technologieën de wijze waarop bedrijven concurreren.

Deze eigenschappen leiden tot een aantal theoretisch aantoonbare verschijnselen, zoals het bestaan van meerdere evenwichtssituaties. Vaak hangt het van kleine toevalligheden af welk evenwicht wordt bereikt. Zo zijn er veelal twee evenwichtssituaties: één waarbij niemand de technologie gebruikt, waardoor de technologie onvoldoende waarde heeft voor een potentiële gebruiker; en een tweede evenwicht waarin vrijwel iedereen de technologie gebruikt, en de technologie daardoor heel waardevol is voor iedere gebruiker. Omdat veelal het economische nut van beide situaties verschilt, is het goed mogelijk dat een economisch systeem in een suboptimaal evenwicht 'gevangen zit'. Een dergelijke situatie wordt aangeduid met de term 'lock-in'. Dit heeft tot de nodige controverse geleid, vooral omdat 'lock-in' moeilijk empirisch aantoonbaar is. Zou de elektrische auto een superieure technologie zijn geweest als er evenveel onderzoeks- en ontwikkelingsenergie in was gestopt als in de fossiele brandstof auto?

De bijdrage van dit proefschrift aan bovenstaand debat is drieledig: (1) er wordt een empririsch voorbeeld van lock-in gegeven, met meetbare inefficiency; (2) de rol van landsgrenzen wordt expliciet beschouwd, deze blijkt een doorslaggevende rol te spelen in de invoering en gelijkschakeling van standaarden en netwerk technologieën; en (3) de opvolging van netwerk technologieën wordt expliciet gemodelleerd. Daarmee worden een aantal bekende eigenschappen van technologische innovatie verklaard.

1. Empririsch voorbeeld van lock-in. Er bestaan grote verschillen tussen landen in het gebruik van betalingsinstrumenten. Zo gebruiken de Verenigde Staten voornamelijk cheques voor het verrichten van betalingen terwijl inwoners van veel Europese landen girale overschrijvingen gebruiken. Diverse auteurs hebben op basis van gedetailleerd kostenonderzoek becijferd dat de Verenigde Staten een kleine 200 miljard dollar goedkoper uit zouden zijn indien ze in plaats van cheques andere instrumenten zoals overboekingen zouden

gebruiken. Hiermee is in ieder geval inefficiency aangetoond, maar is er ook sprake van lock-in? Van veel betalingsverkeer-instrumenten, zoals bijvoorbeeld girale overboekingcircuits, is empirisch aangetoond dat ze onderhevig zijn aan netwerk effecten. Dit proefschrift introduceert een simpel model ter analyse van de beslissing van banken om dergelijke betalingsinstrumenten aan hun cliëntèle te introduceren. Daaruit blijkt dat: (1) er een zekere kritische massa van banken nodig is om de technologie rendabel te introduceren; en (2) het vrijwel nooit loont om binnen een land op betalingsstandaarden te concurreren. De belangrijkste parameters van het model worden gekwantificeerd aan de hand van de historische introductie van giraal betalingsverkeer in Nederland. Introductie blijkt mogelijk indien de daaraan meewerkende banken samen ongeveer de helft van de markt bedienen. Het lijkt dan ook geen toeval dat de introductie van de BankGiroCentrale in 1967 plaatsvond net nadat twee grote fusies voor een aanzienlijke concentratie van het bankwezen zorgden: in 1964 ontstond de ABN uit de Twentse Handels Maatschappij en de Hollandsche Bank Unie, terwijl de Amsterdamse en Rotterdamse Bank fuseerden tot AMRO. Deze twee leidende banken waren samen in staat de overige handelsbanken, en de landbouw banken (Raiffeisen en Boerenleen banken, de latere Rabo) tot actie te bewegen. Samen hadden ze een marktaandeel van zo'n 55%. Anderzijds was het bankwezen in de Verenigde Staten tot 10 jaar geleden zeer gefragmenteerd. Na de concentratiegolf van de afgelopen jaren zijn de banken daar nu actie aan het ondernemen om de kosten van cheque verkeer te verlagen, en giraal verkeer op grotere schaal te introduceren. Het chequegebruik in de Verenigde Staten is daarmee een voorbeeld van lock-in in een inefficiënte technologie.

2. *Landsgrenzen spelen een cruciale rol.* In veel sectoren, zoals betalingsverkeer, post en telecommunicatie, zijn transactiepatronen zeer lokaal. De hoeveelheid grensoverschrijdende transacties is nooit groter dan 5% en vaak rond de 1%. Dit heeft een aantal belangrijke consequenties op de keuze voor netwerkstandaarden. De invloed van buitenlandse standaarden op spelers in een bepaald land is beperkt. Indien de binnenlands standaard afwijkt van die in andere landen heeft dit alleen invloed op de kleine hoeveelheid grensoverschrijdende transacties. Bedrijven binnen een land kunne hierdoor relatief makkelijk een standaard invoeren, zonder zich al te veel te bekommeren om het buitenland. De keerzijde van deze medaille is dat hierdoor elk land al snel een eigen standaard kiest: er is geen dwingende economische noodzaak voor een gemeenschappelijke standaard. De natuurlijke neiging om in elk land het wiel uit te vinden krijgt dan al snel de overhand. En indien de standaarden eenmaal per land verschillen, is het moeilijk en kostbaar om alsnog tot harmonisatie te komen. Moeilijk, omdat coördinatie tussen vrijwel deelnemers (bedrijven) noodzakelijk is. Kostbaar, omdat alle gebruikers op een nieuwe standaard moeten over-

gaan, hetgeen migratiekosten met zich meebrengt die al gauw hoger zijn dan de baten; de directe baten hangen immers af van de beperkte hoeveelheid grensoverschrijdende transacties. Daarnaast zijn er natuurlijk indirecte baten, zoals betere internationale marktwerking. Deze baten vallen echter veelal niet toe aan de betrokken bedrijven. Integendeel, aangetoond kan worden dat internationale standaardisatie leidt tot verhevigde concurrentie en lagere winsten voor de betrokken bedrijven (de keerzijde van de eerder genoemde indirecte baten voor een samenleving). In dat geval hebben bedrijven juist belang bij het handhaven van nationale standaarden.

3. Technologie opvolging en innovatie volgen veelal nationale patronen. Analyse van introductie van nieuwe betaalsystemen in de VS en Nederland wijst uit dat nieuwe netwerk technologieën veelal gebruik maken van bestaande infrastructuren. Zo lift het PIN systeem voor winkel betalingen mee op de PIN-pas die de consument reeds bezat voor gebruik in geldautomaten (ATMs). Dit mechanisme is niet verwonderlijk. Het is niet eenvoudig om nieuwe netwerk technologieën te introduceren, vanwege de kip-ei problemen: het PIN-product is pas aantrekkelijk voor winkeliers indien voldoende consumenten een pas hebben, en omgekeerd is gebruik van de pas voor consumenten pas interessant als voldoende winkeliers hem accepteren. Indien deze cirkel kan worden doorbroken door mee te liften op bestaande infrastructuren maakt dit de zaken aanzienlijk makkelijker. Het leidt er echter ook toe dat de bestaande basis een belangrijke invloed speelt bij de invoering van nieuwe technologieën. Modellering van dit mechanisme leidt tot een aantal belangrijke bevindingen. In de eerste plaats blijken innovatiepatronen vaak nationaal: verschillen in technologie tussen landen drukken hun stempel op de keuze voor nieuwe netwerk technologieën. In de tweede plaats leidt het mechanisme tot 'de wet van de remmende voorsprong' van Jan Romein. Enerzijds zorgt bestaande infrastructuur voor een voordeel vanwege het meelift effect. Anderzijds is er minder ruimte voor nieuwe, betere, technologieën omdat veel van de baten al worden geleverd door bestaande technologieën, zij het tegen hogere kosten. Modelmatige exercities wijzen uit dat dit tweede mechanisme de overhand heeft.

Hoofdstuk 1 beschrijft het betalingsverkeer landschap, en constateert dat er grote en blijvende verschillen tussen landen zijn. Deze verschillen hebben economische consequenties. Zo zijn er diverse bronnen die schatten dat het Amerikaanse chequegebruik tot extra kosten leidt, ter waarde van 0.5-1% van het nationaal product. Tevens analyseert hoofdstuk 1 de historische introductie van betaalsystemen in de VS en Nederland, en constateert het eerder genoemde mechanisme waarbij vaak bestaande infrastructuur elementen worden benut.

Hoofdstuk 2 geeft een overzicht van de bestaande literatuur op 3 terreinen: betalingsverkeer, netwerk technologieën en betalingssystemen als netwerken.

Geconstateerd wordt dat: (1) er een groot aantal inzichten en modellen voor netwerk technologieën bestaat, maar dat de rol van landsgrenzen onderbelicht is, (2) vrijwel alle betaalsystemen aan netwerk effecten onderhevig zijn, en (3) dit de enige plausibele verklaring vormt voor de verschillen tussen landen in de structuur van het betalingsverkeer.

Hoofdstuk 3 en 4 introduceren modellen om de invoering en harmonisatie van standaarden te analyseren. Hoofdstuk 3 doet dit voor open standaarden die voor iedereen toegankelijk zijn, hoofdstuk 4 doet dit voor 'sponsored' standaarden, waarbij bedrijven zelf de toegang kunnen bepalen. De modellen leiden tot de eerste twee van de eerder genoemde drie bijdragen van dit proefschrift.

Hoofdstuk 5 en 6 illustreren de werking van de modellen aan de hand van twee praktijkvoorbeelden: de invoering van giraal betalingsverkeer in Nederland, en de poging tot harmonisatie van het Europese girale betalingsverkeer om te komen tot een Single Euro Payments Area (SEPA). De parameters van de modellen in hoofdstuk 3 en 4 worden geschat, en geconstateerd wordt dat de uitkomsten van de modellen in lijn zijn met de werkelijke uitkomsten.

Hoofdstuk 7 tot en met 9 kijken naar de rol van innovatie en technologie opvolging. Hoofdstuk 7 beschouwt de relevante literatuur, die geen goed model voor de opvolging van (netwerk) technologieën oplevert. Hoofdstuk 8 introduceert een dergelijk model en voert een aantal computer simulaties uit. Dit levert het derde van de eerder beschreven drie bijdragen op. Hoofdstuk 9 illustreert het innovatie mechanisme aan de hand van de invoering van systemen voor betaling via Internet en mobiele telefoon. Van de 200 nieuwe systemen die in de jaren negentig op de markt kwamen zijn er slechts een paar over. Het leeuwendeel van de betalingen via Internet vindt plaats via de reeds bestaande systemen van credit card (VS) en girale betaling (Europa). Dit bevestigt nogmaals de belangrijke rol van bestaande infrastructuur en het feit dat landen hun eigen weg volgen, ook al zijn dezelfde nieuwe technologieën voor iedereen beschikbaar.

Hoofdstuk 10 vat de conclusies samen en hoofdstuk 11 geeft een nabeschouwing. Daarin wordt geopperd dat de mechanismen en bevindingen van dit proefschrift breder toepasbaar zijn. In de Westerse economieën is het meeste kapitaal niet fysiek maar 'intangible'. En netwerk infrastructuren vormen hiervan een belangrijk deel. Nog verdergaand kan men speculeren dat veel van wat 'cultuur' genoemd wordt in feite (gedrags-) standaarden zijn, geschreven (wet- en regelgeving) of ongeschreven. Ook hierop zouden de modeluitkomsten (blijvend nationale patronen en de wet van de remmende voorsprong) van toepassing kunnen zijn.

Curriculum Vitae

Gottfried Leibbrandt (Amsterdam, 24 March 1961) went to secondary school at Gymnasium Sorghvliet, The Hague. He studied mathematics and econometrics at the Free University of Amsterdam (1979-1985) and holds an MBA from Stanford Business School (1985-1987). He worked as a research assistant in Amsterdam (ESI, 1984), and as a summer intern in London (Bain, 1986). Since 1987 he is a management consultant with McKinsey&Company, specializing in retail banking and payments. He has served clients throughout Europe, including several major banks and payment organizations.

Universiteit Maastricht